THE UNIVERSITY *of* EDINBURGH

Epigenetic Regulation of Germline-Specific Genes

Jamie A. Hackett

Presented for the degree of Doctor of Philosophy
University of Edinburgh
January 2010

Declaration

I declare:

(i) that this thesis has been composed by myself, and

(ii) that the work presented here is my own, unless otherwise stated.

Jamie Hackett

January 2010

Acknowledgements

One of the most satisfying aspects of completing this thesis is the opportunity to thank all those who have supported, guided and contributed to this work. In this respect I owe a great deal of gratitude to Richard, who is always immensely supportive, encouraging and a treadmill of ideas. More than this though, Richard, you have allowed me to develop (with gentle hints) and you have been a friend and a mentor. I am indebted to you.

I would also like to thank all the members of the Meehan lab over the last 4 years for help, discussions and most impressively, for putting up with me. A special thanks is reserved for Donncha for his advice, technical know-how and friendship over the years. The heady combination of office banter, 'Friday chat', and lab-golf always kept things lively. I am heartily obliged to have had you around (but also scarred from your horrifying selection of 'music' for the lab). Thanks also to the Masters students James and Emma for their hard toil during their work with me. Beyond the lab, thanks to everyone in CGE for being so accommodating and helpful in all matters science, particularly Ian for help with all things developmental. Thanks to the affectionately known 'langers' (Luke, Donncha, Ross, Seamus) and the 'hen-night' posse (Catherine, Shelagh, Liz, Diana, Rachel) for keeping up the 'craic'. And thanks to all at the HGU who have helped me out, propped me up or shared a chuckle, that I have thoughtlessly forgot to mention.

This thesis would never have happened without the unwavering support and care of my family and especially my parents. Mum and Dad, I owe my deepest gratitude to you for the unquestioning love, nourishment and opportunities in life you have given me and I am forever indebted. This is for you.

Lastly, to my partner in crime, Jen. For upping roots and carting up to Edinburgh with me, for always being there, and for all the good times, I can't thank you enough.

You make smile and you make me happy, you have made this PhD and our lives a wonderful experience. I love you.

Abstract

In mammals, epigenetic modifications and trans-acting effectors coordinate gene expression during development and impose transcriptional memories that define specific cell lineages and cell-types. Methylation at CpG dinucleotides is an epigenetic mechanism through which transcriptional silencing is established and heritably maintained through development. Functionally, DNA methylation regulates key biological processes such as X-chromosome inactivation, transposon repression and genomic imprinting. However, the extent to which DNA methylation is the primary regulator of single-copy gene expression and the precise mechanism of methylation-dependent silencing remain undetermined. Here, I identify a novel set of germline-specific candidate genes putatively regulated by DNA methylation. Analysis of one candidate gene, *Tex19,* demonstrates that promoter CpG methylation is the primary and exclusive mechanism for regulating developmental silencing in somatic lineages. Genetic or pharmacological removal of CpG methylation triggers robust de-repression of *Tex19* and loss of transcriptional memory. Moreover, *Tex19* critically relies on *de novo* methylation, mediated by Dnmt3b, to impose silencing in differentiating ES cells and somatic cells *in vivo* from embryonic day (E)7.5. Reporter gene and ChIP analysis demonstrate that *Tex19* is strongly activated by general transcription factors and is not marked by repressive histone modifications in somatic lineages, consistent with differential DNA methylation *per se* being the primary mechanism of regulating expression. Full transcriptional silencing of *Tex19* is critically dependent on the methyl-binding protein (MBP) Kaiso, which is only recruited to methylated *Tex19* promoter. The reliance on DNA methylation and Kaiso for silencing in somatic cells establishes an epigenetic memory responsible for maintaining expression in germline and pluripotent cell types through successive developmental cycles. This thesis represents the first *causal* report of lineage-specific promoter DNA methylation directing silencing of an *in vivo* gene through recruitment of an MBP.

Table of Contents

List of Figures

List of Tables

Abbreviations

5-Aza dC	5-aza deoxycytodine
5mC	5'methyl cytosine
°C	degrees centigrade
Ac	acetate
ATP	adenosine triphosphate
BCIP	5-Bromo-4-Chloro-3'-Indolyphosphate p-Toluidine Salt
BMP	bone morphogenetic protein
bp	base pairs (of DNA)
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
C. elegans	*Caenorhabtidis elegans*
ChIP	chromatin immunoprecipitation
CpG	cytosine and guanine seperated by a phosphate
C-terminal	carboxy-terminal
cRNA	complementary ribonucleic acid
Da	daltons
dH$_2$0	distilled water
DAPI	4,6-diaminophenylindole
DMEM	dulbecco's modified eagle medium
DMSO	dimethyl sulphoxide
DMR	differentailly methylated region
DNA	deoxyribonucleic Acid
DNase	deoxyribonuclease
Dnmt	dna methylatransferase
dNTP	deoxynucleotide triphosphate
DP	*dnmt1*$^{-/-}$ *p53*$^{-/-}$ MEFs
dpp	days post partum
Drosophila	*Drosophila melanogaster*
DTT	dithiothreitol
E	embryonic day
EB	embryoid body
EC	embryonal carcinoma (cell)
E. coli	*Escherichia coli*
EDTA	ethylenediaminotetra-acetic acid
EG	embryonic germ (cell)
EpiSC	epiblast stem cell
ES	embryonic stem (cell)

EtBr	2,7-diamino-10-ethyl-9-phenyl-phenathridium
EtOH	ethanol
FCS	foetal calf serum
FGF	fibroblast growth factor
g	relative centrifugal force
GFP	green fluorescent protein
GMEM	glasgow's modified eagle medium
GS	germline stem (cell)
GST	glutathione-S-transferase
H3K4	histine H3 lysine 4
H3K9	histone H3 lysine 9
H3K27	histone H3 lysine 27
HAT	histone acetyltransferase
HCl	hydrochloric acid
HDAC	histone deacetylase
HMT	histone methylatransferase
HP1	heterochromatin associated protein 1
Hox	homeobox
ICM	inner cell mass
kb	kilobase pairs of DNA
kDa	kiloDaltons
LB	Luria-Bertani
LINE	long interspersed nucler element
LIF	leukaemia inhibitory factor
LTR	long terminal repeat
M	molar
MBD	methyl-CpG binding domain
MBP	methyl-CpG binding protein
MEF	mouse embryonic fibroblast
mES	mouse embryonic stem (cell)
miRNA	micro RNA
mRNA	messenger RNA
Mw	molecular weight
N2a	Nuero 2a (cells)
NBT	nitro-Blue Tetrazolium Chloride
NP-40	nonidet NP-40
nt	nucleotide
N-terminal	amino-terminal
NTP	nucleotide triphosphate
Oligo	oligonucleotide

ORF	open reading frame
P	*p53*$^{-/-}$ MEFs
PAGE	polyacrylamide gel electrophoresis
PBS(t)	phosphate buffered saline (plus 0.2% w/v Tween-20)
PcG	polycomb group proteins
PCR	polymerase chain reaction
pen	penicillin
PFA	paraformaldehyde
PGC	primordial germ cell
piRNA	piwi-interacting RNA
PMSF	phenyl methyl sulfonyl fluoride
PRC	polycomb repressive complex
qRT-PCR	quantitative reverse-transcriptase PCR
RNA	ribonucleic acid
RT-PCR	reverse transcriptase polymerase chain reaction
rpm	revolutions per minute
RT	room temperature
S. cerevisiae	*Saccharomyces cerevisiae*
SDS	sodium dodecyl sulphate
SINE	short interspersed nuclear element
strep	streptomycin
TAE	tris-EDTA acetic acid buffer
TBE	tris-EDTA boric acid buffer
TBS(t)	tris buffered saline (plus 0.05% w/v Tween-20)
TE	trophectoderm
TEMED	N, N, N', N-tetramethylethylene diamine
Tex19	*Tex19*.1
TF	transcription factor
TSA	trichstatin A
TSS	transcription start site
UTR	untranslated region
UV	ultraviolet
vol	volumes
v/v	volume/volume
w/v	weight/volume
wt	wild-type
X-Gal	5-Bromo-4-Chloro-3-indolyl B-D-galactopyranoside
Xenopus	*Xenopus laevis*

Happy is he who gets to know the reasons for things.

Virgil (70-19BC) Roman poet

Chapter 1

Introduction

1.1 Epigenetics

Conrad H Waddington originally coined the term 'Epigenetics' in 1942 as a portmanteau of epigenesis (the development of an organism) and genetics (the study of hereditary and variation) (Waddington, 1942). To Waddington, epigenetics described a conceptual model of how genotypes might interact with the environment during development to generate a phenotype. Since then epigenetics has been a term of evolving definition. In 1987 Robin Holliday described epigenetics as "the study of changes in gene expression, which occur in organisms with differentiated cells, and the mitotic inheritance of given pattern of gene expression" (Holiday, 1987; Holiday, 1994). Holiday's definition extended Waddington's ideas by acknowledging epigenetic processes are not just restricted to development and also defined the molecular basis as heritable regulation of gene expression. More recently however, Riggs and colleagues narrowed the definition of epigenetics to "the study of mitotically and/or meiotically heritable changes in gene function that cannot be explained by changes in DNA sequence" (Riggs et al., 1996). This definition makes it clear what epigenetics is not - inheritance of mutational DNA change, but leaves open the question of what mechanisms are at work. Because the extent to which many epigenetic processes, such as histone modifications, are truly heritable is uncertain (although it is apparent they are at least re-imposed) Riggs and colleague's definition has been refined to its current incarnation. Thus, in the modern sense, epigenetics is the study of *stable* changes in gene expression potential that are not attributable to alterations in DNA (Bird, 2007). In the broadest sense, any additional information stably superimposed onto the DNA sequence can be considered 'epigenetic'.

Epigenetic phenomena explain several deviations away from normal Mendelian genetics including paramutation, position effect variegation (PEV) and transgene inactivation (Coe Jr, 1968; Karpen & Allshire, 1997; Chandler, 2007). These examples challenge the contemporary view that phenotype is dictated linearly by DNA and suggest that DNA sequence alone is insufficient to determine the observed phenotype. Perhaps the most striking role of epigenetics though, is in directing and maintaining the differentiation of a totipotent zygote to the myriad cell-types of a multicellular organism (Gurdon, 1992; reviewed by Hemberger et al., 2009). Here, the progressive restriction of potential that characterises development is guided by coordinated changes in gene expression. The correct spatial and temporal expression of these genes is achieved by epigenetic processes and trans-acting factors that impose stable transcriptional memories that define the lineage and developmental potential of each cell. Thus, cellular differentiation can be considered an epigenetic process, governed by changes in what Waddington described as the 'epigenetic landscape' (*Fig 1.1*) (Waddington, 1957). Epigenetics is in short, a collection of mechanisms that considerably extends the information potential of the genetic code through manipulation, but not alteration, of the genotype. To accomplish this, epigenetic processes influence chromatin structure and accessibility (Wolffe & Matzke, 1999).

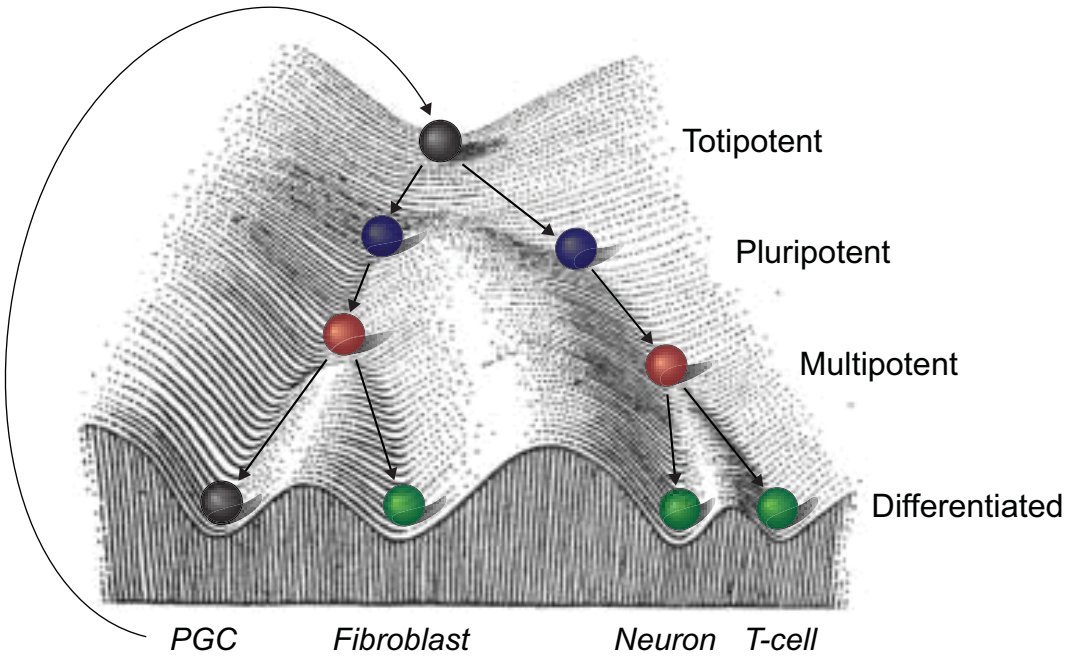


Fig 1.1. **Epigenetic restriction of developmental potential - Waddington's canal.** An adaptation of C. H. Waddington's epigenetic landscape, showing cell populations with different developmental potentials. Cellular development is illustrated as marbles rolling down a landscape (epigenetic restriction) into one of several potential valleys (cell fates). Colored marbles correspond to various differentiation states. Black, totipotent; Navy, pluripotent; Red, multipotent; Green, unipotent. PGCs (shown in black) represent a unique specialised cell type, that upon fertilisation, can acquire totipotent epigenetic and developmental potential. Differentiated cells can also be experimentally reprogrammed to pluripotent iPS cells by defined factors (Takahashi & Yamanaka, 2006). Figure modified from Waddington (1957).

1.2 Chromatin Structure

Eukaryotic DNA is packaged with histone and non-histone proteins into chromatin. This has two important consequences. Firstly, it allows the packaging and compaction of ~1.8m of DNA into a nucleus typically 5-20μm in diameter. Secondly, it promotes the regulation of essential cellular events including transcription, lineage specification, DNA replication and cell division. The fundamental unit of chromatin is the nucleosome, which is composed of an octamer of the four core histones (H2A, H2B, H3 and H4) around which 147bp of DNA is super-helically coiled 1.65 times (Corner & Thomas, 1974; Allan et al., 1980; Luger et al., 1997). Additionally linker histones, which do not form part of the core nucleosome, bind to the structure through their interaction with both DNA and core histone H2A (Boulikas et al., 1980). Each nucleosome is coupled via 10-70bp of linker DNA producing an ~11nm chromatin fibre, the so-called 'beads on a string' arrangement (Olins & Olins, 1974). This structure can further compact to a 30nm chromatin fibre, which is the functional unit of chromatin. Despite its importance, the structure of the 30nm fibre remains undetermined, with evidence to support both the bent linker DNA 'solenoid' and straight linker DNA 'zigzag' models (Finch & Klug, 1976; Woodcock et al., 1984; Robinson & Rhodes, 2006). Ultimately chromatin is able to resolve to highly condensed mitotic chromosomes but the mechanistic basis of this 'higher order' compaction is unclear (Khorasanizadeh, 2004).

Grossly, there are two types of chromatin environment – transcriptionally silent heterochromatin and active or at least permissive, euchromatin, each of which is associated with characteristic epigenetic modifications. Heterochromatin can be further sub-divided to constitutive and facultative forms (Brown, 1966). Constitutive heterochromatin (c-Het) is invariant between cells and during development and mostly contains repetitive centromeric, telomeric and satellite sequences. When stained with 4,6-diamino-phenylindole (DAPI) c-Het is seen as dark foci (particularly in mice) due to the condensed structure of the chromatin and because DAPI preferentially binds the AT-rich sequences that compose the c-Het satellite

repeats. Epigenetically, c-Het is marked by hypermethylated CpG dinucleotides, hypoacetylated histones, and tri-methylated histone 3 lysine 9 (H3K9me3) which recruits heterochromatin protein 1 (HP1) (Lachner et al., 2001; Kouzarides, 2007). The favoured model predicts these modifications interact to form a positive feedback loop that stably maintains chromatin condensation. Here, HP1 bound to H3K9me3 via its chromodomain, acts as docking platform to recruit other repressive factors to heterochromatin (Li et al., 2002a; Smallwood et al., 2007; Bergman & Cedar, 2009). Unlike c-Het, facultative heterochromatin (f-Het) represents sub-regions of the genome where the chromatin compaction state changes throughout development and can therefore be considered dynamically silenced. In contrast, euchromatin is an 'open' structure that is either permissive or actively transcribed and marked by hyperacetylation and di- and tri- methylation of histone 3 lysine 4 (H3K4me2/3). Mechanistically, the decondensed nucelosome arrays in euchromatin allow regulatory proteins access to bind DNA and function.

Within the broad domains of euchromatin and heterochromatin are complex landscapes of epigenetically modified histones and DNA including punctate sites at regulatory elements, insulators and transcriptional start sites (Barski et al., 2007). These localised epigenetic modifications guide the gene-specific expression patterns required for a cell to differentiate and remain stably lineage-committed. The sum total of all epigenetic information in each cell is termed the 'epigenome' (Goldberg et al., 2007). Because the epigenome is highly variable between cell types and over time there must be at least as many epigenomes, and hence chromatin structures, as there are cell types at any given time-point. The importance of the epigenome to complex mammalian development is epitomised by the so-called C-value enigma (Gregory, 2001; Gregory, 2002). This points out that while the roundworm *C.elegans* is a considerably simpler organism than humans and correspondingly its genome is approximately 30 times smaller, both organisms contain ~20,000 genes. Therefore, enhanced organismal complexity and genome size are not accompanied by an increase in gene number. This presents the problem that there are a vastly amplified number of cryptic binding sequences for transcription factors not associated with genes, which could lead to deregulated and ectopic gene expression. To address this,

it is suggested that epigenetic systems have co-evolved with increasing genome size to organise chromatin into permissive gene-containing and inaccessible closed portions (Wray et al., 2003). For example, the invertebrate to vertebrate transition was paralleled by the emergence of global DNA methylation to stably silence the transcriptional noise that was an inescapable by-product of enhanced genome complexity (Bird & Tweedie, 1995). This process has led to the intricate epigenetic regulation that promotes complex mammalian development.

1.3 Epigenetic Mechanisms in Mammals

Many epigenetic mechanisms contribute to establishing and maintaining the transcriptional profile of a mammalian cell. Of these, methylation of CpG dinucleotides and modification of histones are the best characterised (Jaenisch & Bird, 2003). However, nucleosome variants, sub-nuclear localisation, replication timing and regulatory non-coding RNAs also contribute to the epigenetic landscape of a cell (Mahy et al., 2002; Donaldson, 2005; Fraser & Bickmore, 2007). Broadly, these epigenetic mechanisms regulate transcription by directly influencing chromatin structure or being interpreted by proteins that recognize a precise epigenetic signal and affect the appropriate biological response (Klose & Bird, 2006). These systems may operate independently or, more commonly, functionally interact to form feedback loops that perpetuate gene expression states. Within epigenetic feedback loops different mechanisms may have subtly different roles. For example, one epigenetic modification may initiate a gene expression state while another recruited by the initiating mark, stably maintains that state. Because of this combinatorial regulation by epigenetic processes, each mechanism discussed here cannot always be considered in isolation.

1.3.1 Histone Modifications

In 1964, Allfrey and colleagues demonstrated that the core histone proteins are subject to multiple post-translational modifications and speculated this might have a role in gene activity (Allfrey et al., 1964). Since then over 100 modifications of defined residues have been characterised, largely at the N-terminal tails that protrude from nucleosomes, but also in the core globular domains (*Fig 1.2*) (Bernstein et al., 2007; Mersefelder & Parthun, 2006). The major effect of histone modifications is to regulate transcription by creating a docking site for non-histone effecter proteins to modify chromatin structure. However in some cases, strongly charged modifications can alter chromatin structure directly, through disrupting the DNA-histone interaction (Lachner et al., 2001; McGhee & Felsenfeld, 1980; Turner, 2000). A diverse range of histone modifications have been reported including; acetylation,

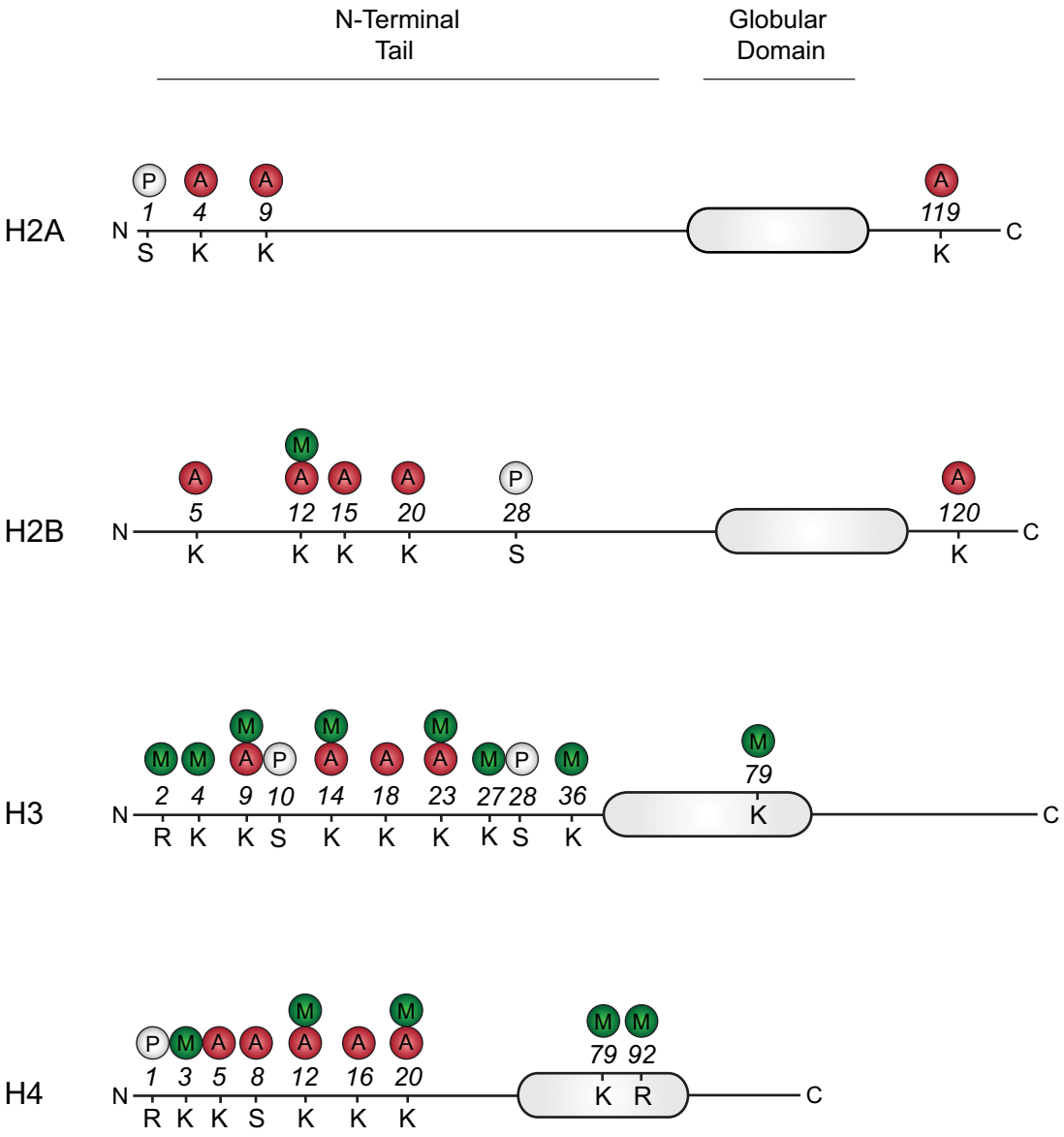


Fig 1.2. **Epigenetic modifications of core histones.** Defined residues in each histone protein can be modified by A: acetylation (red), M: methylation (green) and P: phosphorylation (white). Different epigenetic and transcriptional consequences have been observed depending on the specific residue modified and the modification itself. The amino acid residues of each of the four core histones are shown using standard single letter nomeclature. K: lysine, R: arginine , S: serine. The location of each residue within the primary sequence of the histone is shown by the italicised number. Only a subset of the reported modifications is shown.

methylation, phosphorylation, ubiquitination, sumoylation, formylation, deimination, ADP ribosylation, and proline isomerization (Kouzarides, 2007). Each of these modifications is generally associated with a particular transcriptional outcome depending on the chemical modification itself, the histone residue modified and, the site within the transcriptional unit. The effect on local transcription of some of the well characterised histone modifications and variants are shown in *Table 1.1*.

Histone Modification/Variant	Typical Location	Transcriptional Effect	Reference
Ac H3/H4	TSS	++	Roh et al., 2006
H3K4me1	Bimodal over TSS	+	Barski et al., 2007
H3K4me2	Bimodal over TSS	++	Kim et al., 2005
H3K4me3	TSS	++	Kim et al., 2005
H3K9me1	TSS	+	Barski et al., 2007
H3K9me2	Promoter	-	Bannister et al., 2001
H3K9me3	Promoter	-	Mikkelsen et al., 2007
H3K27me1	Promoter	+	Barski et al., 2007
H3K27me2	Promoter	-	Boyer et al., 2006
H3K27me3	Promoter	-	Boyer et al., 2006
H3K36me1	Gene body	+	Barski et al., 2007
H3K36me3	Gene body	++	Joshi & Struhl, 2005
H3K79me1	TSS	-/+	Barski et al., 2007
H3K79me2	TSS	-/+	Barski et al., 2007
H3K79me3	TSS	+	Vakoc et al., 2006
H4K20me1	Gene body	++	Talasz et al., 2005
H4K20me3	TSS	-	Schotta et al., 2004
H3R2me1	Promoter	-/+	Barski et al., 2007
H3R2me2 (as)	Promoter	-/+	Barski et al., 2007
H4R3me1	TSS	-/+	Huang et al., 2005
H2BUb1	Gene body	+	Zhu et al., 2005
H2AUB1	TSS	-	Wang et al., 2006
H2A.Z	Bimodal over TSS	++	Bruce et al., 2005
Macro-H2A	TSS	-	
H3.3	TSS	++	Chow et al., 2005

***Table 1.1.* Effect of histone modifications on transcription.** Key, ++ is strongly positively correlated with transcription, + is positively correlated with transcription, -/+ shows no observable correlation and – is negatively correlated with transcription. TSS: transcriptional start site.

1.3.1.1 The histone code

The great potential for different combinations of modifications at any particular locus can introduce a large degree of variation into the chromatin template. This

forms the basis of the 'histone code', which postulates that multiple modifications provide a cumulative or sequential code that dictates the local chromatin landscape (Strahl & Allis, 2000). Ultimately, this predicts that we will be able to "read" the functional status of any region through the combination of histone modifications present. In an alternative model, Bernstein and Schreiber argue that multiple modifications can act as a signalling network to confer 'bistability, robustness and adaptability" (Schreiber & Bernstein, 2002). Recently, several genome-wide studies have examined the distribution of histone modifications across the genome in an effort to 'crack' the histone code. These have used the availability of high-throughput sequencing technology to give nucleotide level resolution of histone modifications in the context of transcriptional status and developmental stage (Birney at al., 2007; Barski et al., 2007; Mikkelsen et al., 2007).

1.3.1.2 Histone Methylation

The most prevalent modification of core histones is methylation, which can occur in the mono-, di- or tri- methylated forms at lysine residues or symmetrical and asymmetrical forms at arginines (Bannister et al., 2002; Zhang & Reinberg, 2002). These modifications are dynamically deposited and removed by groups of enzymes collectively referred to as histone methyltransferases (HMTs) and histone demethylases (HDMs), respectively. Unlike acetylation and phosphorylation, it is thought that histone methylation does not directly alter chromatin structure but recruits downstream protein complexes, which recognise methylated residues through chromo-domains (Kouzarides, 2007).

Early evidence suggested methylation of histone 3 lysine 4 (H3K4) was associated with the transcriptionally active macronucleus of *Tetrahymena* (Strahl et al., 1999). Subsequent mammalian genome wide and localised studies have confirmed that H3K4 mono-, di- and tri- methylations all associate with active promoters albeit each with different distributions around the transcriptional start sites (TSS) (*Table 1.1*) (Bernstein et al., 2002; Barski et al., 2007; Heintzman et al., 2007). Interestingly, work in the erythroid lineage showed that H3K4me2 is always present at promoters

marked by H3K4me3 but that the reciprocal is not true, suggesting H3K4me2 is required for H3K4me3 deposition (Orford et al., 2008). The derivatives of H3K4 can be used to discriminate different types of regulatory element, with mono- and di-methylated H3K4 marking dynamic enhancers and di- and tri- H3K4 present at active promoters (Heintzmann et al., 2007). Recently, it was shown that H3K4me3 at active promoters is recognised by the TAF3 component of TFIID, directly linking this histone modification to the transcriptional machinery, in addition to its role in recruiting co-activators (Vermeulen et al., 2007).

Methylation of H3 Lysine 9 (H3K9) is generally regarded as a repressive modification. H3K9me3 is highly enriched at heterochromatic regions, where it creates a binding site for heterochromatin protein 1 (HP1) (α, β & γ isoforms), and appears to be the default state for inactive portions of the genome (Rea et al., 2000; Lachner et al., 2001; Bernstein et al., 2007). In contrast, H3K9me2 appears to dynamically mark repressed regions in a tissue specific manner. Feldman (2006) and colleagues have reported that G9a catalysed H3K9me2 is the initiating mark that creates a local heterochromatin structure at the proximal promoter of the pluripotency factor *Oct3/4* during ES cell differentiation. Moreover, highly conserved 'blocks' (up to 5Mb) of H3K9me2 modified chromatin accumulate in differentiated cells and are correlated with repression of local genes (Wen et al., 2009). Surprisingly, H3K9me1 has been found to be more closely associated with active transcriptional start sites (Barski et al., 2007).

Histone H3 lysine 27 (H3K27) methylation was first described as a repressive modification catalysed by the PRC2 complex (Cao et al., 2002). Subsequent genome-wide studies have confirmed H3K27me3 associates with repressed genes, particularly key developmental loci, including the *Hox* clusters (Boyer et al., 2006; Lee et al., 2006). Intriguingly, Bernstein and colleagues (2006) noted that in ES cells many repressed regions marked by H3K27me3 were also modified with H3K4me3, a mark normally associated with gene activity. Sequential ChIP analysis confirmed that the same allele carried both the repressive and activating modification, leading to the term 'bivalent domain'. Following differentiation, the authors found that

bivalency generally resolved to either the H3K4me3 or H3K27me3 states, leading to either activation or repression of the locus (Bernstein et al., 2006). However, more recent studies have also identified bivalent domains in progenitor and differentiated cells (Mohn et al., 2008; Barski et al., 2007; Mikkelsen et al., 2007). Because many bivalently marked loci are key developmental genes, it is postulated that the bivalent state keeps genes repressed but 'poised' or 'primed' to rapidly respond to downstream developmental cues, without the requirement for epigenetic reprogramming at the loci.

1.3.1.3 Histone Acetylation

Acetylation of histones was one of the first modifications to be discovered and is now known to be required for nucleosome assembly and to promote transcriptional activity (Allfrey, 1964; Mello & Almouzni, 2001; Roh et al., 2005). The relationship between histone acetylation at promoters and gene activity was initially described through chromatin immunoprecipitation (ChIP) studies (Hebbes et al., 1988). This association has been supported by genome-wide ChIP analysis using microarrays (ChIP on chip) and mass sequencing (ChIP-seq) (Roh et al., 2005; Mikkelsen et al., 2007; Heintzman et al., 2007). Mechanistically, acetylated histones neutralize the charge interaction between positive histone tails and the negative DNA backbone, which decondenses chromatin leading to a permissive state for active transcription (McGhee & Felsenfeld, 1980). In addition, proteins containing a bromo-domain can recognise and bind to acetylated histones and mediate downstream effects (Dhalluin et al., 1999). *In vivo*, the acetylation state of histones is rapidly maintained in a dynamic equilibrium by enzymes that catalyse acetylation, histone acetyltranferases (HATs), and remove the modification, histone deactylases (HDACs). The dynamic interaction of these two processes is essential, as inhibition of HDACs results in severely disrupted development of amphibian and mammalian embryos and mis-expression of many genes in mammalian cells (Almouzni et al., 1994; Lagger et al., 2002; Lande-Diner et al., 2007).

There are many other reported modifications of histones *in vivo*, including methylation of lysines 36 and 79 of histone H3, lysine 20 of H4 in addition to methylation of arginine residues (Hublitz et al., 2009). Other modifications such as ubiquitination of H2A and H2B, sumoylation of all four core histones and proline isomerization have also been observed (Wang et al., 2005; Nathan et al., 2006; Nelson et al., 2006). However while genome-wide studies are developing an understanding of the function of these marks, not a great deal is known about their ability to modulate transcription and chromatin structure or indeed other processes such as recombination, repair and replication (Kouzarides, 2007).

1.3.1.4 Crosstalk

The presence or absence of a particular histone mark does not always lead to a specific transcriptional outcome. This is at least partly because many histone modifications 'crosstalk' to establish expression states. For example, H3K9me3 recruits HP1 to promote a repressive chromatin structure. However, modification of the adjacent serine residue (H3S10) by phosphorylation prevents HP1 interacting with H3K9me3, leading to a reduction in heterochromatin bound HP1 despite the presence of its binding platform, H3K9me3. Functionally, transient phosphorylation of H3S10 may occur during mitosis to allow replication factors access to heterochromatin while maintaining a memory for which regions of the genome should be silenced. Consistent with this and as a further example of histone crosstalk, in *S. cerevisiae* phosphorylation of histone H3S10 promotes acetylation of H3 lysine 14 establishing an 'open' chromatin template (Walter et al., 2008). Thus in keeping with the histone code model, each histone modification and/or variant should not always be thought of as an independent regulatory mark, but as part of a complex interconnected mechanism.

1.3.2 Core Histone Variants

With the exception of Histone H4, non-allelic variants of all the core histones have been identified, which have roles in gene regulation, DNA repair and chromosome segregation (Malik & Henikoff, 2003; Pursala & Bhargava, 2005). The important

cellular role of histone variants is highlighted by the fact that some variants show a greater debris of cross-species conservation than the canonical histones.

1.3.2.1 Variants of H2A

Of the variants of H2A, Macro-H2A is incorporated into heterochromatin and key developmental genes in ES cells where it may promote transcriptional repression while H2ABdb has a mutually exclusive pattern indicating it may have a role in active regions (Bushbeck et al., 2009; Costanzi & Pehrson, 1998; Chadwick & Willard, 2001). H2A.X has a specialised role in localising to sites of double stranded DNA breaks and targeting them for repair (Malik & Henikoff, 2003). The widely studied H2A.Z seems to have the biggest role in gene regulation and is essential for both *Drosophila* and mammals (Clarkson et al., 1999; Faast et al., 2001). In yeast HA.Z is associated with active gene promoters and chromatin domain boundaries, with similar patterns observed in metazoans (Meneghini et al., 2003; Farris et al., 2005). Consistent with a role in transcriptional activation and separation of functional chromatin domains, H2A.Z was found to be mutually exclusive of DNA methylation in *Arabidopsis* (Zilberman et al., 2008). The incorporation of H2A.Z at promoters is thought to modify the nucleosome surface structure, which aids in the recruitment of transcription factors and co-activators (Raisner & Madhani, 2006).

1.3.2.2 Variants of H3

The H3 variant H3.3 is a replication-independent histone that, like H2A.Z, accumulates at actively transcribed promoters and has been observed to incorporate into a transgene array *in vivo* as it undergoes activation (Chow et al., 2005; Janicki et al., 2004). Incorporation of H3.3 is thought to destabilise the nucleosome, allowing access for transcription factors. Interestingly, the combination of H3.3 and H2A.Z creates the least stable nucleosome, suggesting these variants could functionally interact (Jin and Felsenfeld, 2007). In *Xenopus,* Histone H3.3 has been suggested to promote epigenetic memory by maintaining an active chromatin structure following replication (Ng & Gurdon, 2008). The replication-dependent histone H3 variants, H3.1 and H3.2 differ by one amino acid and are commonly considered to be

functionally equivalent, although a recent report has suggested H3.2 may be associated with transcriptionally repressed chromatin (Hake et al., 2006). CENP-A is the final H3 variant and has a highly specialised and essential role in chromosomal kinetochore formation (Palmer et al., 1991).

1.4 DNA Methylation

As early as 1975 DNA methylation was proposed to be an epigenetic system responsible for maintenance of a gene expression state through mitotic divisions (Holliday & Pugh, 1975; Riggs, 1975). CpG methylation has since been recognised as an important contributor to the stability of gene expression patterns through promoting the formation of transcriptionally inactive chromatin in diverse organisms (Wolffe & Matzke, 1999). Importantly, DNA methylation satisfies Riggs (1996) stringent definition of an epigenetic system in that it is truly mitotically and meiotically heritable. In prokaryotes, both adenine and cytosine residues are found to be methylated as part of the host restriction system that selectively cleaves unmethylated foreign DNA. However, in multicellular eukaryotes, methylation of DNA only occurs at cytosine residues that reside predominately, but not exclusively, within the genomic context of a CpG dinucleotide (Hotchkiss, 1948; Wyatt, 1950, Ramsahoye et al., 2000; Bestor, 2000). Interestingly, a novel modification of DNA, 5'hydroxymethylation has recently been reported in mouse neurons and ES cells, but the functional significance and distribution of this modification is yet to be determined (Kriaucionis & Heintz, 2009; Tahiliani et al., 2009). Likewise, the precise role of non-CpG methylation, which can represent up to 25% of total cytosine methylation in ES cells, is yet to be established (Ramsahoye et al., 2000, Lister et al., 2009).

CpG methylation plays an essential role in vertebrate development and consequently high levels of DNA methylation are found in vertebrate genomes (Li et al., 1992; Suzuki & Bird, 2008). In contrast, many invertebrates including *D.melanogaster* and *C.elegans* have low or (almost) undetectable genomic DNA methylation levels and the yeasts *S.cerevisiae* and *S.Pombe* also lack DNA methylation (Tweedie et al., 1997; Proffitt et al., 1984; Antequera et al., 1984). This appears to suggest that DNA methylation *per se* is not an evolutionarily requisite for development. However, increasingly, genomic methylation is being found to play important roles in organisms where it had previously been overlooked. For example, cytosine methylation has recently been shown to regulate retrotransposon silencing and

telomere integrity in *Drosophila* (Phalke et al., 2009). Moreover many invertebrates *do* exhibit significant global levels of DNA methylation (Field et al., 2004). In the honeybee *A.Melifera,* genomic methylation plays a crucial role in social structure organisation. Here, under normal conditions larvae nurtured with royal jelly develop into queens whereas unnurtured larvae of the same clonal origin mature into worker bees. Inhibition of *de novo* methylation induces larvae to develop into queens irrespective of exposure to royal jelly, suggesting DNA methylation epigenetically stabilises key developmental specifications (worker or queen) in honeybees (Kucharski et al., 2008; Barchuk et al., 2007). DNA methylation is also essential for normal development in non-mammalian vertebrates such as *Xenopus laevis* and zebrafish *Danio rerio* (Stancheva & Meehan, 2000; Anderson et al., 2009). However, as the focus of this thesis is primarily on mammalian systems, the discussion herein concentrates on DNA methylation in the context of mammals.

In mammals, it is proposed that the fundamental role of DNA methylation is to repress transcription and heritably maintain a silenced chromatin state (Bird, 1984). This function has been directly demonstrated in several distinct biological processes including X-chromosome inactivation, imprinting, cell differentiation, repression of transposable elements and stable reinforcement of epigenetically silenced genes (Heard et al., 1997; Li et al., 1993; Jackson et al., 2004; Walsh et al., 1998; Feng et al., 2006; Bird & Wolffe, 1999). Because CpG methylation is involved in many essential cellular functions, aberrant patterns of methylation are implicated in diverse pathologies such as carcinogenesis, disease susceptibility and in aging (Bird, 2002; Egger et al., 2004). Furthermore, genetic studies with mice lacking DNA methyltransferases (Dnmts) have demonstrated DNA methylation is essential for embryonic development and for survival of differentiated cells (Li et al., 1992; Okano et al., 1999, Jackson-Grusby et al., 2001; Damelin & Bestor, 2007). Despite the biological importance of DNA methylation, many questions regarding its function remain, including; how is it targeted or excluded from specific regions of the genome, what is the mechanism through which transcriptional repression is achieved, and what influences do the complex interactions with other epigenetic systems have.

1.4.2 Distribution of DNA methylation across the mammalian genome

1.4.2.1 CpG islands

In differentiated cells, approximately 4% of cytosines are methylated which corresponds to ~80% of all potential CpGs (Gruenbaum, 1981). However, the distribution of DNA methylation across the mammalian genome is not uniform, but bimodal. Almost all CpGs are methylated except those that reside within discrete fractions of the genome termed CpG islands (CGIs), which remain largely unmethylated. CGIs thus represent short regions of DNA enriched in CpGs but devoid of methylation, at least in the germline, which ensures they do not undergo the mutational loss of CpGs (through deamination) that afflicts the rest of the genome (Bird, 1986, Suzuki & Bird, 2008). In mammals, CGIs have an average length of ~1,000bp, are GC-rich compared with bulk genomic DNA and are typically associated with gene promoters or regulatory elements. Indeed, approximately ~60% of mouse genes including most housekeeping genes have CGI promoters and, consistent with this association, CGIs are generally characterised by a transcriptionally permissive chromatin state (Tazi & Bird, 1990; Zhu et al., 2008; Weber et al., 2007; Guenther et al., 2007).

Identification of CGIs is generally achieved computationally using a criteria of observed to expected ratio (Obs/Exp) of 0.6 and >55% GC content over a 500bp window (Takai & Jones, 2002). The Obs/Exp ratio indicates the fraction of CpGs present compared to what would be expected if this dinucleotide were randomly distributed throughout the genome. However, because CpG dinucleotides have been evolutionarily depleted from mammalian genomes, at most loci this ratio is generally very low (0.1-0.2). In some studies the length parameter has been reduced to 200bp to enrich the number of CGIs, but this can result in up to a 10-fold excess of false positives, largely Alu elements. Recently, a method of CGI detection based on sequencing human blood fragments has been developed that takes into account clustering but uniquely, also requires the absence of methylated CpGs (Illingworth et al., 2008). Mechanistically, it is hypothesised that CGIs remain unmethylated due to their interactions with active transcription factors or through characteristic histone

modifications such as H3K4me2, which may preclude the methylation machinery from binding to CGIs directly or indirectly (Weber et al., 2007; Meissner et al., 2008). In support, the methyltransferase cofactor Dnmt3L only recognizes CpGs associated with unmodified H3K4 and recruits and/or stimulates Dnmt3a mediated *de novo* methylation at these sites (Ooi et al., 2007; Jia et al., 2007). As all hypomethylated CpG-rich promoters are constitutively H3K4 methylated, regardless of activity, this modification may account for the hypomethylated state of CGIs (Barrera et al., 2008; Mohn et al., 2008; Guenther et al., 2007). Alternatively, Antequera & Bird (1999) proposed that CGIs may be replication origins and their unmethylated status represents a 'footprint' as a consequence of this function. However, the precise mechanism that excludes most CGIs from genomic methylation remains an open question (Illingworth & Bird, 2009).

1.4.2.2 Promoter methylation

Several recent studies have used massive parallel bisulphate sequencing or methylated DNA ImmunoPrecipitation (MeDIP) to examine the DNA methylation landscape at promoters in the context of gene activity. A study in humans sub-divided promoters to distinguish strong CpG islands (HCP), weak CpG islands (ICP) and sequences with no local enrichment of CpGs (LCP) (*Fig 1.3a*). According to these classifications, most autosomal HCPs (97%) were hypomethylated, regardless of transcriptional status whereas the majority of LCPs were methylated, but the low density of CpGs was not sufficient for gene repression (*Fig 1.3*) (Weber et al., 2007). On the rare occasions when HCPs were methylated, they were silenced, indicating that repression by DNA methylation requires a high local density of 5mC. Interestingly, ICPs were distinct in that although they were generally unmethylated, they were preferential targets for *de novo* methylation in somatic cells compared to the germline and this was sufficient to silence transcription. This indicates that weak CpG islands are more prone to methylation as development proceeds, and potentially suggests a role for *de novo* methylation in repressing lineage-specific genes associated with ICPs. (Weber et al., 2007). In support, a second study in mice also observed that weak CpG islands (ICPs) were associated with *de novo* methylation

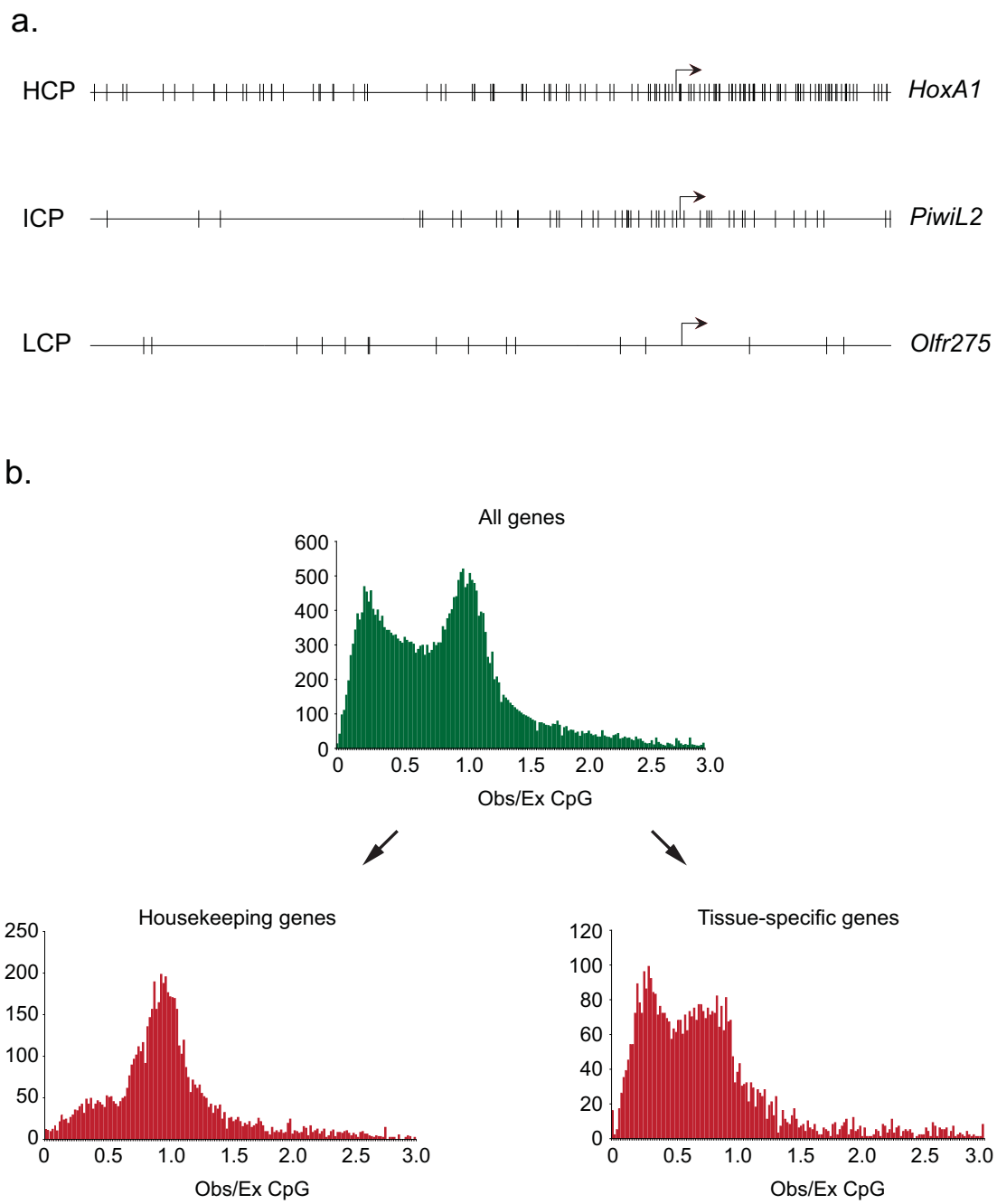


Fig 1. 3. **Distribution of CpG dinucleotides in mammalian genomes.** a.) Examples of the promoter classifications: low CpG-density (LCP) (*Olfr275*), intermediate CpG-density (ICP) (*PiwiL2*) and high CpG-density (HCP) (*HoxA1*). The horizontal line represents the region -2000bp to +600bp of the transcriptional start siite (TSS) which is represented by an arrow. Each vertical line represents a single CpG dinucleotide. b.) A histogram of gene promoter CpG ratio (observed/expected) in the mouse genome according to tissue of expression. Gene expression data for 48 mouse tissues was retrieved from the BioGPS database and CpG content was determined for all promoters in the dataset (n = 13, 729). Total genes exhibit a bimodal distribution. However, ubiquitously expressed housekeeping genes (expression level>150 in >30 tissues) are mostly CpG-rich indicating these genes are not regulated by promoter CpG methylation. Tissue-specific genes (expression level>150 in <10tissues) show almost equal numbers of CpG-rich and -poor promoters. Concept from Mohn & Schubeler, 2009.

during differentiation and following extended proliferation *in vitro* (Meissner et al., 2008). These observations suggest that DNA methylation is not a general mechanism for dynamically regulating the expression of genes associated with HCPs and LCPs either tissue-specifically or developmentally. In contrast, genes associated with ICPs, particularly germline specific genes, frequently show differential methylation patterns that strongly correlate with gene activity, indicating weak CpG island promoters are preferential targets for regulation by DNA methylation. Whether the susceptibility of ICPs to dynamic DNA methylation reflects shared sequence features, the characteristics of the larger chromatin domain, or an intrinsic property is a question of great interest. Recently, an intriguing study has suggested that DNA methylation may be the only epigenetic mark at 30% of silenced genes in mouse ES cells. Interestingly, the pool of genes where DNA methylation was the only mark are greatly enriched in ICPs (Fouse et al., 2008). This indicates that DNA methylation could function exclusively of other epigenetic systems at some promoters, particularly weak CpG islands, while complementing other epigenetic regulatory mechanisms at others.

1.4.2.3 Global Methylation patterns

Several large-scale studies have recently investigated the genomic DNA methylation pattern outside of gene promoters. A mass-sequencing study of 1.9 million CpG sites on human chromosomes 6, 20 and 22 in twelve tissues confirmed the bimodal nature of the genomic landscape, with the majority of genomic regions being heavily methylated and CpG-rich regulatory sequences being hypomethylated (Eckhardt et al., 2006). Interestingly, this report and others also noted that consistent with other non-CGI regions of the genome, gene-bodies were hypermethylated, even when transcriptionally active (Rabinowicz et al., 2003; Rakyan et al., 2004). It is hypothesised that this gene-body methylation may have a role in preventing aberrant initiation and transcriptional interference outside of the canonical TSS. Indeed global DNA methylation *per se* may function to inhibit cryptic transcriptional initiation (Suzuki & Bird, 2008; Weber & Schubeler, 2007). Eckhardt and colleagues also noted many tissue-specific differentially methylated regions (T-DMR) that were

correlated with conserved non-coding sequences but very few genomic regions that were differentially methylated according to age or gender. In contrast to this later result a recent study reported that dizygotic twins had significantly discordant DNA methylation patterns compared with monozygotic twins (Kaminsky et al., 2009). Other global analyses have suggested the mammalian genome contains many T-DMRs. A study using restriction landmark genomic sequencing (RLGS) with *NotI,* identified 247 T-DMRs that varied between embryonic stem (ES), embryonic germ (EG) and trophoblast stem (TS) cells (Shiota et al., 2002). However, the functional significance of most T-DMRs has yet to be tested and, notably they rarely occur in promoter CGIs where the CpG density is high enough to influence transcription.

1.4.3 Enzymes that deposit DNA Methylation in mammals

The enzymes responsible for targeting and maintaining global DNA methylation are known as DNA methyltransferases (Dnmt). In mammals, they are constructed from a complex set of functional modules, broadly divided into the N-terminal 'regulatory' domain and the C-terminal 'catalytic' domain. Dnmt2, which lacks a regulatory domain, is the exception to this rule. The regulatory domain functions largely as an interaction module, allowing multiple protein-protein interactions, DNA binding and nuclear localisation (Chen & Li, 2006). Conversely, the C-terminal domain comprises ten motifs responsible for the enzyme's catalytic activity. Six of these motifs are conserved in nearly all cytosine methyltransferases across the evolutionary spectrum from bacteria to mammals (*Fig 1.4*) (Posfai et al., 1989). Indeed, DNA methyltransferases in general are highly conserved among eukaryotes with methylated genomes.

1.4.3.1 de novo vs maintenance methyltransferases

At least three methyltransferase enzymes, Dnmt1, Dnmt3a and Dnmt3b coordinate the establishment and maintenance of DNA methylation patterns in mammals (Li, 2002a; Goll & Bestor, 2005). These enzymes have conventionally been organised into two general classes according to their preferred substrate. The '*de novo'* methyltransferases, Dnmt3a and Dnmt3b, target cytosine methylation to previously

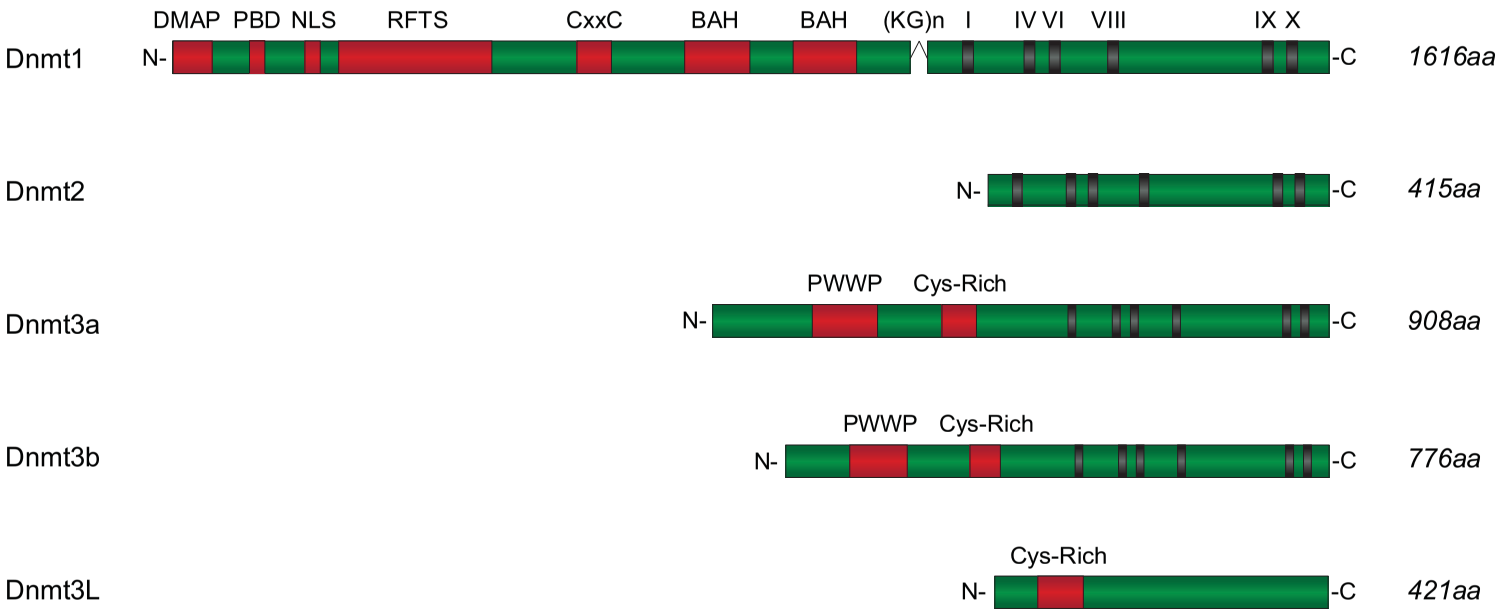


Fig 1.4. **Structure and organisation of mouse Dnmt protein family.** Schematic diagram of the maintenance methyltransferase (Dnmt1), the *de novo* methyltransferases (Dnmt3a & Dnmt3b) a catalytic co-factor (Dnmt3L) and a putative RNA methyltransferase (Dnmt2). Key domains are shown in red. The N-terminal domain of Dnmt1 contains a proliferating cell nuclear antigen-binding domain (PBD) that contributes to targeting to the replication fork, a nuclear localization signal (NLS), an CXXC DNA-binding motif, a bromo-adjacent homology domain (BAH), and replication foci targeting sequence (RFTS). Dnmt3a/b exhibit a tetrapeptide PWWP motif essential for binding to chromatin and a cysteine-rich (cys-rich) domain for protein interactions. The C-terminal catalytic domains includs six conserved motifs shown in black: I is involved in the formation of the AdoMet binding site; IV binds the substrate cytosine at the active site; VI contains the glutamyl residue serving as a proton donor; VIII's function is unclear; IX maintains the structure of the substrate-binding site; and X participates in the formation of the AdoMet binding site. AdoMet - S-adenosylmethionine; aa – amino acids.

unmethylated CpG dinucleotides, whereas the '*maintenance'* enzyme, Dnmt1, preserves existing methylated sites. (Okano et al., 1999; Bestor et al., 1988; Siedlecki & Zielenkiewicz, 2006). Despite these classifications, the distinction between *de novo* and maintenance methyltransferases is not clear-cut. For example, the effect of disruption of *Dnmt3a* and *Dnmt3b* in ES cells suggests that these enzymes have an essential role in maintaining global DNA methylation. Here, an extended absence (70 passages) of Dnmt3a and Dnmt3b renders the genome severely hypomethylated, with only 0.6% of CpGs retaining methylation as opposed to ~70% in wild-type ES cells (Jackson et al., 2004; Chen et al., 2003). Moreover *in vitro*, the 'maintenance' methyltransferase Dnmt1 has greater *de novo* methylation activity on unmethylated substrates than Dnmt3a and Dnmt3b, and has recently been implicated in *de novo* methylation of CpG islands *in vivo* (Jair et al., 2006; Bestor et al., 1988; Esteve et al., 2005a). Indeed, Dnmt1 can interact with Dnmt3a and Dnmt3b in a co-operative complex *in vivo*, blurring the distinction between *de novo* and maintenance methylation activity (Kim et al., 2002). Despite this, the consensus remains that the primary role for Dnmt1 is in maintaining methylation patterns, whereas that of Dnmt3a and Dnmt3b is in *de novo* synthesis of CpG methylation.

1.4.3.2 Dnmt1

The first eukaryotic cytosine methyltransferase to be purified and cloned was Dnmt1 (Bestor et al., 1988). A 5-30 fold preference for hemi-methylated CpGs over unmethylated substrates led to its characterisation as the maintenance methyltransferase, whereby it duplicates pre-existing patterns of methylation during replication. This classification was additionally based on Dnmt1's interaction with replication foci during S-phase, which promotes efficient methylation of newly synthesised strands but is not strictly required (Chuang et al., 1997; Jeltsch, 2006; Schermelleh et al., 2007). By preserving patterns of methylation at each replication cycle, Dnmt1 provides heritability to genomic methylation allowing long-term potentiation of epigenetic states (Yoder et al., 1997).

Dnmt1 is essential for normal vertebrate development - *Dnmt1* deficient mice die around embryonic day (E)8.5, at the onset of gastrulation (*Table 1.2*) (Li et al., 1992). As such, the biological role of Dnmt1 has been investigated largely through genetic studies in murine embryonic stem (ES) cells and embryonic fibroblasts (MEFs). Disruption of *Dnmt1* by homologous recombination generates several novel phenotypes (*Table 1.2*). The most striking consequence of *Dnmt1* deletion is a 95% reduction in genomic methylation levels (Lei et al., 1996). Other reported effects include a 10-fold increase in the mutation rates of exogenous marker genes in ES cells and bi-allelic expression of several, but not all, imprinted genes (Chen et al., 1998; Li et al., 1993). For example, *H19* and *Kcnq1ot1* are expressed from both alleles whereas *Igf2* and *Kcnq1* are expressed from neither (Goll & Bestor, 2005). Partial inactivation of both X-chromosomes has been observed in mutant embryos due to the aberrant cis-activation of *Xist* (Panning and Jaenisch., 1996). Counter-intuitively, there is also upregulation of many X-linked genes in ES cells, as a result of the demethylation of their associated CpG island promoters (Fouse et al., 2008). Finally, loss of *Dnmt1* results in extensive demethylation of LTR regions and consequentially, expression of IAP transposable elements (Walsh et al., 1998). The phenotypic severity of transposon deregulation has led many to postulate Dnmt1 primarily evolved to repress transposons by maintaining methylation at these sequences (Goll and Bestor, 2005). However, neither Dnmt1 deficient cells nor hypomethylated cancer cells have shown enhanced rates of transposition (Zilberman et al., 2007; Suzuki & Bird, 2008).

An important and as yet not fully understood finding is that ES cells deficient in *Dnmt1* develop normally in an undifferentiated state but die of p53-mediated apoptosis when induced to differentiate (Li et al., 1992). Interestingly, Dnmt1-deficient ES cells can initiate differentiation, as shown by alkaline-phosphotase negative colonies, but lack the capacity to differentiate down specific lineages (Jackson et al., 2004). The same p53-mediated apoptotic fate is exhibited by mouse embryonic fibroblasts (MEFs) when *Dnmt1* is conditionally deleted *ex vivo* (Jackson-Grusby et al., 2003). Apoptosis is also the underlying cause of death in *Xenopus* embryos depleted of *Dnmt1* (Stancheva et al., 2001).

Protein	Functions	Mutant Phenotype	Reference
Methyltransferases			
Dnmt1	Maintenance methylation; putative roles in DNA repair and replication and direct repression	Embryonic lethal by E8.5; hypomethylation; loss of imprinting and X-linked gene expression; transposon activation. ES cells Viable	Li et al, 1992 Lei et al, 1996
Dnmt1o	Oocyte-specific isoform	Loss of maternal imprints	Howell et al, 2001
Dnmt2	Non-CpG methylation (Drosophila); RNA methylation	No phenotype	Okano et al, 1998
Dnmt3a	*De novo* establishment of methylation	Postnatal lethal (~4weeks); Spermatogenesis defects, defective *de novo* methylation	Okano et al, 1999
Dnmt3b	*De novo* establishment of methylation	Embryonic lethal by E14-18, defective *de novo* methylation of minor satellite; ICF syndrome	Okano et al, 1999
Dnmt3L	Co-factor required for maternal methylation	Abnormal maternal imprinting; male sterility	Bourc'his et al, 2001 Hata et al, 2002
Methyl-binding proteins			
MeCP2	Methyl-binding protein; recruit co-repressors	Mild neurological abnormalities in adults	Guy et al, 2001
Mbd1	Methyl-binding protein; recruit co-repressors	Behaviour abnormalities; Defects in adult neurogenesis	Zhao et al, 2003
Mbd2	Methyl-binding protein; recruit co-repressors	Defects in maternal behaviour	Hendrich et al, 2001
Mbd3	NuRD complex	Embryonic lethal by E6.5	Hendrich et al, 2001
Mbd4	Repair enzyme, methyl-binding protein; recruit HDACs	Increased mutation frequency	Millar et al, 2002
Kaiso	Sequence specific & methyl-binding protein; recruit co-repressors	Resistance to intestinal cancer	Prokhortchouk et al, 2006
Zbtb4	Methyl-binding protein	ND	
Zbtb38	Methyl-binding protein	ND	
Histone-modifying proteins			
Eed	Histone H3 lys27 methylation	Embryonic lethal	Niswander et al, 1988
Hdac1	Histone deactylation	Embryonic lethal by E10.5, excessive histone acetylation	Lagger et al, 2002
G9a	Histone H3 lys9 di-methylation	Embryonic lethal by E8.5; loss of K3K9 methylation in euchromatin	Tachibana et al, 2002
Suvar39h1 Suvar39h2	Histone H3 lys9 tri-methylation	Embryonic lethal by E14.5, chromosomal instability, increased tumour risk	Peters et al, 2001
Ezh2	Histone H3 lys27 tri-methylation; PRC2/3 complex	Peri-implanation lethal; defect in primitive ectoderm	O'Carroll et al, 2001

Table 1.2. **Mouse mutant phenotypes of selected epigenetic mediators.**

Somatic Dnmt1 is a ~190kDa protein expressed strongly in proliferating cells and ubiquitously in somatic tissues throughout mammalian development (Chen & Li., 2006) but is present at significantly lower levels in noncycling cells. Throughout G1-phase, Dnmt1 is diffusely distributed around the nucleoplasm but co-localises with replication foci during S-phase when its cell cycle regulated expression is highest. Dnmt1 is transcribed from 3 promoters and utilizes 5' sex-specific exons. An oocyte specific form, Dnmt1o (175kDa), is transcribed from the most 5' promoter and lacks the N-terminal 118 amino acids of somatic Dnmt1. The truncation stabilizes Dnmt1o and allows the variant to accumulate to high levels in the cytoplasm of noncycling mouse oocytes via an interaction with annexin V (Ding & Chaillet, 2002; Doherty et al., 2002). The functional significance of this is currently unclear, although some evidence suggests that Dnmt1o provides maintenance methylation specifically at imprinted loci during the fourth embryonic S-phase (Howell et al., 2001). The second and primary promoter, 6kb 3' of the oocyte specific site, produces the full-length Dnmt1 transcript expressed in somatic and ES cells. The third *Dnmt1* promoter, which is the most 3', is used during sperm development and produces a non-translated transcript (Dnmt1s) that may be utilized to down-regulate Dnmt1 protein levels in pachytene spermatocytes (Mertineit et al., 1998).

Structurally, the large N-terminal regulatory domain of Dnmt1 is separated from the catalytic domain by 13 alternating glycyl and lycyl residues. The N-terminal portion harbours several functional domains including a nuclear localisation sequence (NLS), DMAP interaction module, a PCNA-binding domain (PBD) and a bromo adjacent homology domain (BAH). Moreover, the N-terminal module is responsible for targeting Dnmt1 to hemi-methylated DNA during replication. The PBD domain interacts with proliferating cell nuclear antigen (PCNA) during S-phase, which concentrates Dnmt1 activity at replication forks (Chuang., 1997; Hermann et al., 2004). This interaction increases the efficiency of maintenance methylation but is not required for the inherent processivity of Dnmt1 (Vilkaistis et al., 2005; Schermelleh et al., 2007). The intrinsic sequence specificity for CpG, but not hemi-methyl CpG, is a property of the catalytic domain as demonstrated by domain swap experiments.

Here, the catalytic domains of prokaryotic methyltransferases with diverse target sequences were fused to the N-terminal region of Dnmt1. This did not alter the target specificities of the prokaryotic catalytic domains but did impart a novel preference for hemi-methylated DNA (Pradhan & Roberts, 2000). An additional cysteine rich zinc finger DNA binding domain (CXXC) is located centrally within the N-terminus and appears to facilitate unmethylated CpG DNA binding (Fatemi et al., 2001).

Dnmt1 has been reported to interact with an increasingly large number of proteins including; DMAP1 (Dnmt1 associated protein), Rb, E2F1 transcription factor, hSNF2H (ATP-dependent remodelling factor), PCNA, Suvar39 (histone methyltransferase), HDAC1 and 2 (histone deacetylase 1 and 2) and NP95 (Chuang et al., 1997; Rountree et al., 2000; Robertson et al., 2000; Fuks et al., 2003 Goll & Bestor, 2005; Sharif et al., 2007). The wealth of interactions between Dnmt1 and chromatin modifying proteins provides a link between DNA methylation, histone modifications and chromatin structure, which serves to stably reinforce epigenetic states. Additionally, the reported interactions with several key co-repressors support a role for Dnmt1 as part of a direct transcriptional repression complex. Dnmt1 has been reported to form complexes with HDAC1, Rb and E2F that repress transcription independent of changes in DNA methylation (Fuks et al.,2000; Robertson et al., 2000). Furthermore, knock down of Dnmt1 induces activation of unmethylated Sp1-responsive genes in mice (Milutinovic et al., 2004). A recent report in human cancer cells has also noted general mitotic catastrophe when Dnmt1 is completely depleted. Importantly the effect was observed before replication associated DNA demethylation could occur, indicating Dnmt1 has an essential function aside from maintenance methylation (Chen et al., 2007). In *Xenopus*, a partial knock down of xDnmt1 results in ectopic gene activation independently of changes in promoter methylation. Here, gene silencing could be re-imposed by both wt and crucially, catalytically mutant forms of DNMT1, strongly supporting an *in vivo* non-catalytic repressive function for Dnmt1 (Dunican et al., 2008). Further studies have shown that xDnmt1 *per se* may also function as part of a DNA damage recognition complex and downstream apoptotic inducer (Ruzov et al., 2009c). However, in the context of ES cells, the catalytic function of Dnmt1 is essential to

rescue the differentiation phenotype of *Dnmt1*-null cells (Damelin & Bestor, 2007). Thus, the general significance of direct repression by Dnmt1 in mammalian systems is yet to be determined (*Supplementary thesis chapter available, see pp 273*).

1.4.3.3 Dnmt3a & Dnmt3b

Two functional cytosine methyltransferases reside in this family, Dnmt3a and Dnmt3b. Expression of these enzymes is greatest during early developmental stages, where their primary role is genome-wide *de novo* methylation, although they show the same propensity for hemi-methylated DNA (Okano et al., 1999). Following gastrulation, Dnmt3a and Dnmt3b activity decreases leading to a profound reduction of *de novo* methylation during differentiation. The expression of Dnmt3b remains very low except in the testes, thyroid and bone marrow whereas Dnmt3a is expressed at reduced levels ubiquitously (Xie et al., 1999). Dnmt3a null mice survive to term but are runted and die within 4 weeks with progressive loss of germ cells in males (*Table 1.2*). However, global DNA methylation patterns seem to remain intact (Okano et al., 1999). Dnmt3b deficient mice die in late gestation (~E14.5) but *Dnmt3b* hypomorphs can survive to adulthood and display some of the characteristics of Immunodeficiency, Centromere instability and Facial anomalies (ICF) syndrome, which is associated with point mutations in *DNMT3B* in humans (*Table 1.2*)(Okano et al., 1999). The more severe phenotype exhibited by *Dnmt3b-/-* mice as compared to *Dnmt3a-/-,* suggests that Dnmt3b has a more important role in embryonic development. Consistent with this, Dnmt3b is expressed earlier during embryonic development, with Dnmt3b *de novo* activity first appearing in the inner cell mass at the implantation stage (Watanabe et al., 2002). Combined deletion of *Dnmt3a* and *Dnmt3b* is embryonic lethal at ~E8.5, equivalent to the *Dnmt1-/-* phenotype but with a lesser extent of demethylation (Okano et al., 1999). The embryonic lethality of both the *Dnmt1*-/- and *Dnmt [3a-/-, 3b-/-]* mice demonstrates that maintenance and *de novo* methylation are both required for development. Double *de novo* mutant ES cells retain methylation in early passages including most imprints, but are progressively demethylated during extended culture (Jackson et al., 2004). This suggests that Dnmt1 is inherently inefficient at maintaining methylation and that in wild-type cells, reiterative *de novo* methylation is required to preserve

methylation patterns. In support of this a recent study found Dnmt1 to be <99% efficient (Schermelleh et al., 2007). Unlike Dnmt1 deficient ES cells, *Dnmt [3a-/-, 3b-/-]* ES cells can differentiate into embryoid bodies, which consist of all three germ layers, but only at early passages (<25), before they are severely hypomethylated (Jackson et al., 2004).

Dnmt3a and Dnmt3b are highly homologous at their C-terminal catalytic domain but differ at their N-termini (*Fig 1.4*). The respective N-terminal domains are responsible for targeting chromatin, via the PWWP domain, and imparting subtle differences in specificity. In this respect, it appears Dnmt3b is specialised for methylation of specific regions of the genome such as pericentromeric repeats and CpG islands on the inactive X, whereas Dnmt3a is required for maternal imprints of differentially methylated regions (DMRs), in addition to their general *de novo* roles (Siedlecki & Zielenkiewicz, 2006). Protein interaction domains in the regulatory N-termini also mediate binding to transcriptional co-repressors (Qiu et al., 2002). Dnmt3b is associated with Sin3a, SUMO1/Ubc9, condensin and the chromatin remodelling enzyme hSNF2H while Dnmt3a has been shown to bind the co-repressor RP58 and the oncogenic factor PML-RAR (Goll & Bestor, 2005). Both also interact with Dnmt1 and histone deacetylases (HDAC). These observations indicate Dnmt3a and Dnmt3b are part of regulatory mechanisms that modulate transcription through recruitment of co-repressors able to influence chromatin structure in addition to *de novo* methylation.

1.4.3.4 Dnmt2

Of the known cytosine methyltransferase homologues, Dnmt2 is the most highly conserved. It is also the most widely distributed homologue across species (Hermann et al., 2004). This strong sequence conservation, particularly within the 10 catalytic motifs, indicates a role in cytosine methylation (Okano et al., 1998; Yoder & Bestor, 1998). Despite this, as of yet Dnmt2 has undetectably weak or no methyltransferase activity in mammals but intriguingly, is the primary enzyme responsible for non-CpG and CpG methylation in early *Drosophila* embryos where it functions to repress

retrotransposons and maintain telomere integrity (Hermann et al., 2003; Phalke et al., 2009). Recent work has demonstrated an additional, potentially novel function in mammals. *In vitro,* purified Dnmt2 can specifically methylate cytosine residues in $tRNA^{Asp}$ but not in DNA (Goll et al., 2006). The purpose of this modification and why it requires such strong conservation of Dnmt2 remain questions of great interest.

1.4.3.5 Facilitators of DNA methylation

An increasing number of non-methyltransferase proteins have been shown to be required to maintain global levels and/or discrete regions of DNA methylation. These 'facilitator' proteins play an essential role in targeting or stimulating methyltransferase activity or in sensitising the chromatin structure to *de novo* DNA methylation. One key facilitator protein, which belongs to the Dnmt3 family, is Dnmt3L. This protein is similar to Dnmt3a and Dnmt3b at its terminal regions but crucially lacks the central PWWP and ATRX domains and is restricted to only mammals. Currently, no methyltransferase activity has been attributed to Dnmt3L and it is not necessary for zygotic development. However, Dnmt3L is specifically expressed in the germ cells where it functions as a regulatory co-factor that stimulates the methyltransferase activity of Dnmt3a and Dmnt3b up to 15-fold (Gowher et al., 2005). This activity is necessary for the establishment of maternal imprinted methylation patterns (Hata et al., 2002; Kaneda et al., 2004). In males, Dnmt3L deficiency results in impaired spermatogenesis probably as a result of the demethylation and aberrant expression of interspersed repeated sequences in germ cells (*Table 1.2*) (Bourc'his & Bestor, 2005; Webster et al., 2005). Deficiency of a second facilitator, Lsh, leads to global demethylation at satellite sequences, imprinted genes and single-copy genes and, perinatal death (Dennis et al., 2001; Xi et al., 2007; Geiman et al., 2001). Lsh has homology to chromatin remodelling proteins and may promote a permissive chromatin structure for methylation. Alternatively, it may stimulate or target *de novo* activity through its interaction with Dnmt3a and Dnmt3b (Zhu et al., 2006). A third protein Uhrf1 (also known as NP95), has recently been shown to be essential for maintaining genome-wide DNA methylation through its interaction with Dnmt1. Uhrf1 preferentially binds hemi

methylated CpGs through its SRA (SET and RING associated) domain and thus targets Dnmt1 to these sites. Disruption of *Uhrf1* leads to a mislocalisation of Dnmt1 in S-phase and a reduction in global CpG methylation levels (Bostick et al.,2007; Sharif et al., 2007).

1.4.4 Reaction Mechanism

The mechanism of DNA methylation was originally proposed by Wu and Santi in 1987 and subsequently modified by Verdine and colleagues (Wu & Santi, 1987; Bestor & Verdine, 1994). A reactive cysteine in the conserved prolycysteinyl motif (PC box) of domain IV initiates nucleophilic attack at carbon-6 (C6) of the pyrimidine ring. This causes activation of carbon-5 (C5) and the transfer of a methyl group from the reaction methyl donor, S-adenonsylmethionine (SAM). Subsequently, the methyltransferase enzyme is released from the 5'-6' dihydro intermediate by β-elimination. Interestingly, a study that used 5-fluocytosine to trap methyltransferases onto their CpG substrate revealed that to gain access to target sequences, the enzyme 'flips' the cytosine out of the double helix during catalysis (Klimasauskas et al., 1994). The method of trapping DNA methyltransferases onto chromatin has also been used to examine the biochemically-induced effects of hypomethylation. Here, the nucleotide analogue, 5-aza deoxycytidine (5-aza dC) can be incorporated into DNA, including at CpG sites. Crucially, 5-aza dC has a nitrogen at position 5 of the pyrimidine ring instead of a carbon. Because DNA methyltransferases are unable to methylate the nitrogen, they cannot release through β-elimination and become covalently trapped onto the DNA resulting in a significant depletion in the amount of enzyme available for methylating normal CpG sequences (Taylor & Jones, 1982). The reduction in the pool of the 'maintenance' methyltransferase, Dnmt1, leads to passive demethylation in cultured cells after two divisions, as synthesis proceeds without methylation of the new strand. Additionally, 5-aza dC may operate through a second mechanism whereby Dnmt1 protein is depleted directly, through targeting for proteosomal degradation (Ghoshal et al., 2005).

1.4.5 Regulation of transcription by DNA methylation

In general, DNA methylation is an epigenetic modification that is responsible for the heritable propagation of gene silencing patterns. Promoter CpG methylation can promote stable silencing even when sequence specific transcription factors are present enabling the formation of transcriptional memory and maintenance of lineage commitment. However, the global distribution of CpG methylation at promoters suggests that expression of only a small proportion of genes can be controlled by DNA methylation (Weber et al., 2007; Meissner et al., 2008). These analyses show CpG dense promoters are constitutively unmethylated regardless of gene activity whereas methylated promoters tend to have CpG densities to low to efficiently silence transcription. Thus, DNA methylation does not seem to be a wide-ranging mechanism for dynamically regulating gene expression. Instead it is largely restricted to the stable silencing of transcription within specific, but crucial, biological contexts such as imprinting and X-chromosome inactivation.

1.4.5.1 X-chromosome Inactivation

In each somatic cell of female mammalian embryos one randomly chosen X-chromosome is transcriptionally silenced (X_i), except prior to the epiblast stage when the paternal X is preferentially silenced (Heard et al. 1997; Mak et al., 2004; Okamoto et al., 2005). This repression is mediated by the expression of the non-coding, but functional RNA *Xist*. The *Xist* RNA silences genes on the chromosome from which it was expressed, by coating the X-chromosome in *cis* and promoting the acquisition of repressive histone marks (Kohlmaier et al., 2004). This appears to be sufficient to silence all the target genes on the inactive X-chromosome. However, many gene promoters, including CGIs, subsequently undergo *de novo* methylation following implantation (Lock et al., 1987; Norris et al., 1991). Promoter methylation seems to contribute to stably maintaining gene silencing, as experimental hypomethylation through 5-aza dC treatment or conditional deletion of *Dnmt1* leads to derepression of genes on the X_i (Czankovski et al., 2001; Sado et al., 2000). Additionally, in extra-embryonic tissues and marsupials, where X-inactivation takes

place without DNA methylation, genes on the X_i become reactivated over time (Samollow et al., 1995; Migeon et al., 1989; Cedar & Bergman, 2009)

1.4.5.2 Imprinting

Imprinted genes are a class of mammalian autosomal genes that are mono-allelically expressed in the embryo and/or adult. The mono-allelic expression of imprinted genes critically relies on parent of origin DNA methylation signals. Generally, one parentally derived allele is extensively methylated while the other exhibits little of no methylation at CpG-rich regulatory sequences (DMRs) (Plass & Soloway, 2002). Imprinted DNA methylation is introduced during gametogenesis, according to sex, by Dnmt3a and Dnmt3L. Following fertilization, imprinted DMRs are preserved by Dnmt1o during cleavage divisions and then by the somatic form of Dnmt1 in embryonic and adult tissues (Howell et al., 2001; Li et al., 1993).

Imprinted methylation affects expression of genes in several, often complex ways, which vary at different imprinted loci. In the simplest cases, DMRs, which overlap the CGI promoter of an imprinted gene, can directly silence expression when methylated. However, imprinting frequently involves the mono-allelic methylation of imprinting control regions (ICRs), which can be many kilobases from the gene or gene clusters they regulate in *cis*. These ICRs can regulate imprinted genes through two broad mechanisms (Reik, 2007). Firstly, they can act as epigenetically modulated chromatin insulators. When unmethylated, these sequences recruit chromatin-insulator proteins that prevent interactions between distant enhancers and promoters, leading to silencing. One such example is the maternally expressed *H19* locus and the nearby paternally expressed *Igf-2* gene. Here, the chromatin-organising protein CTCF binds to the unmethylated maternal ICR, but not the methylated paternal loci, and blocks interactions between *Igf2* and an enhancer 80kb downstream of *H19*. This causes silencing of *Igf2* on the maternal allele and the reciprocal activation of *H19* (Hark et al., 2000; Szabo et al., 2000). In *Dnmt1*-null mice embryos, the paternal *H19* allele is aberrantly expressed whereas the paternal *Igf-2* is aberrantly silenced due to loss of the DNA methylation imprint and CTCF

recruitment (Li et al., 1993; Kato & Sasaki, 2005). The second mechanism involves functional non-coding nuclear RNAs that are regulated by DMRs overlapping their promoter. Expression of imprinted non-coding RNAs from the unmethylated DMR leads to silencing of target genes in *cis* through accumulation of repressive histone modifications (Umlauf et al., 2004; Lewis et al., 2004). Here, silencing may be targeted by the RNA during transcription or alternatively through a localised coating mechanism as observed for *Xist* RNA (Kanduri et al., 2006; Lewis et al., 2006; Reik, 2007).

1.4.5.3 Transposable elements

Mutated interspersed DNA transposons and retro-transposons constitute over 45% of the genome, and represent a severe hazard to the genetic integrity of the organisms they reside in (Smit & Riggs, 1996). One manifestation of this threat is through insertional mutagenesis. Equally, chimeric transcripts and aberrant gene expression mediated by transposon promoters and antisense transcripts, can cause detrimental effects (Goll & Bestor, 2005). In response, mammals utilize DNA methylation as a primary mechanism to stably silence transcription of some transposons, particularly Intercisternal A Particle (IAP) elements. In addition to mediating promoter repression, DNA methylation irreversibly inactivates transposable elements over time through promoting accumulation of cytosine to thymidine transitions via deamination of m^5C (Schorderet & Gartler, 1992). In the absence of DNA methylation, a significant increase in IAP transcription has been reported in early embryos and conditional mutant fibroblasts (Walsh et al., 1998; Jackson-Grusby et al., 2001). Furthermore, in spermatogonia prevented from *de novo* methylating their genome by *Dnmt3L* deletion, LINE and IAP retrotransposons were highly transcribed (Bourc'his et al., 2007; Webster et al., 2005). Aberrant activation of transposable elements has also been reported in globally demethylated Lsh mutant mice (Dennis et al., 2001; Xie et al., 2007). To maintain the heritable repression of transposable elements, many are partially resistant to the erasure of DNA methylation in the zygote and in PGCs. The preservation of methylation at these sequences in the zygote is probably mediated by Dnmt1o and serves to prevent

mutational changes during this epigenetic reprogramming event (Gaudet et al., 2004). How Dnmt1o promotes the maintenance of IAP methylation and indeed imprints, in the context of a progressively demethylated zygotic genome is unclear.

1.4.5.4 Single-copy genes

While there is irrevocable evidence that DNA methylation modulates the activity of imprinted genes, transposons and the maintenance of X-chromosome inactivation, the extent to which promoter methylation regulates developmental and tissue-specific gene expression is unresolved. It is noteworthy that the role of DNA methylation in long-term stable gene silencing is undisputed. However in general, repression by DNA methylation is considered to occur downstream of other epigenetic or trans-acting factors that signal the initial inactivation event. For example, in the case of the pluripotency factor *Oct3/4*, initial repression during differentiation is mediated by sequence-specific repressors such as GCNF (Fuhrmann et al., 2001). The locus is then progressively silenced through acquisition of G9a mediated H3K9 dimethylation and recruitment of HP1. These events lead to localised heterochromatization at the loci and only then, *de novo* methylation by Dnmt3a and Dnmt3b, which functions to reinforce the epigenetic state and prevent aberrant re-expression (Feldman et al., 2006; Chen et al., 2003). Indeed, in differentiating *Dnmt[3a, 3b]* null ES cells, *Oct3/4* can occasionally be reactivated. This indicates that promoter methylation is a secondary system to add an additional layer of repression at the *Oct3/4* loci and in effect, mediates the maintenance rather than initiation of gene silencing (Feldman et al., 2006; Jackson-Grusby et al., 2001; Lande-Diner et al., 2006). Consistent with a maintenance role, a CpG-free transgene is silenced comparably to a CpG-containing transgene but only the CpG containing transgene is resistant to re-activation due to the presence of DNA methylation (Feng et al., 2006).

While the dynamic silencing of *Oct3/4* does not rely on DNA methylation, it may represent a specialised example of a highly regulated gene that employs the combinatorial effects of several epigenetic systems to ensure sustained repression.

This may reflect the vital requirement to maintain *Oct3/4* silencing in order to preserve lineage commitment, prevent de-differentiation and protect against carcinogenesis (Hochedlinger et al., 2005). In contrast, many genes display relatively few levels of epigenetic regulation and these may be candidates for direct regulation by CpG methylation. Indeed, in ES cells ~30% of repressed genes are marked only by DNA methylation and between 5-10% of genes are de-repressed in the absence of global methylation in ES and somatic cells (Fouse et al., 2008; Jackson-Grusby et al., 2001). To distinguish between the genes directly regulated by DNA methylation and those activated as a secondary effect of global mis-expression in these reports, several studies have correlated gene expression *in vivo* to promoter methylation. Here, work on the cell-type specific expression of the *maspin* gene potentially suggests a role for DNA methylation in tissue specific regulation. The promoter of human *maspin* is unmethylated in tissue types that actively express the gene but hypermethylated in tissues in which the gene remains silent (Futscher et al., 2002). Importantly, transcriptional repression is relieved by inhibition of Dnmt1 and loss of promoter methylation.

Work correlating gene expression with DNA methylation has also noted a number of CGI promoter genes that are specifically hypomethylated and expressed in germ cells but repressed and methylated in somatic tissues. These germline genes are often found de-repressed in 5' aza dC treated cells and in hypomethylated tumours, leading to the term cancer/testis (CT) antigens. Most prominent in this group are the MAGE genes, but *Pgk2, Pdha-2* and *Ant4* have also been reported to be demethylated and expressed specifically in germ cells (De Smet et al., 1999; Zhang et al., 1998; Iannello et al., 2000; Rodic et al., 2005). A recent report has also suggested that DNA methylation may mediate the temporal activation of *Dazl, Mvh* and *GCNA* in murine germ cells (Maatouk et al., 2006). Here, the promoters are heavily methylated in somatic cells and primordial germ cells (PGCs) preceding their arrival at the embryonic gonads at E10.5, but are demethylated coincident with initiation of expression in PGCs by E13.5. This work is consistent with global methylation analyses, which have indicated an enrichment of germ cell specific genes associated with hypermethylated HCP or ICP promoters in somatic cells (Weber et al., 2007;

Meissner et al., 2008). Because ICP and HCP promoters are rarely found to be methylated, this suggests that the promoter methylation observed at germline specific genes could represent a specialised function for DNA methylation in somatic silencing of this class of genes.

While an increasing number of genes, particularly germline associated genes, have been reported to be regulated by DNA methylation, it is important to note that currently these associations are largely correlative. The lack of cause and effect reports leaves open the possibility that promoter methylation states at the reported loci could be a consequence of transcriptional activity and not a cause of gene activity *per se*. Alternatively, DNA methylation might maintain repression at these loci, but like at *Oct3/4*, may be a secondary epigenetic mark directed by dynamic histone modifications and not the primary system. To address this, one study has used a phenotypic assay to demonstrate the importance of DNA methylation at the *Elf5* gene. Here, the *Elf5* promoter is hypomethylated and expressed in the trophoblast lineage but methylated and silenced in the epiblast. Methylation and repression of *Elf5* in the epiblast results in fixation of lineage commitment, as shown by *Dnmt1* deficient ES cells, which lose their stable embryonic lineage restriction and are able to adopt trophoblast cell fates *in vitro* and in chimeras (Ng et al., 2008b). Thus, lineage specific DNA methylation at the *Elf5* loci regulates gene expression, as manifested through stable lineage commitment.

Increasingly, the role of DNA methylation in regulating single-copy genes seems to be in restricting expression of key developmental genes and germ cell genes to specific lineages (Hayashi et al., 2008; Ng et al., 2008; Maatouk et al., 2006). By this rationale DNA methylation is essential for the progressive reduction in developmental potency and consistent with this, hypomethylated ES cells proliferate normally until induced to differentiate. Moreover, 5-aza dC treated somatic cells can be reprogrammed to induced pluripotent cells (iPS) at greatly enhanced rates (Mikkelsen et al., 2008). However, the scarcity of cause and effect reports continues to sustain the argument that DNA methylation is a secondary ‘maintenance’ system

for single-copy gene regulation and primarily functions in other processes such as repression of transposons and imprinting.

1.4.6 Mechanisms of Repression by DNA methylation

Two primary models have emerged to describe the mechanism through which DNA methylation at regulatory sequences can influence transcription in mammalian cells. Firstly, methylated CpGs may directly preclude some transcription factors and regulatory proteins from binding to DNA. Depending on the genomic context and nature of the protein excluded from DNA, this can have either positive or inhibitory effects on transcription (*Fig 1.5*) (Watt & Molloy, 1988; Bell & Felsenfeld, 2000). Secondly, DNA methylation can create docking sites for proteins that specifically recognise methylated CpGs. These proteins, known as methyl-CpG binding proteins (MBPs), recruit co-repressors that modulate chromatin structure and modify histones leading to local heterochromatization and transcriptional repression (*Fig 1.5*) (Boyes & Bird, 1991; Jones et al., 1998; Ng et al., 1999). Additionally, methyltransferases *per se* may directly contribute to repression at methylated and non-methylated sequences (*Fig 1.5*) (Rountree et al 2000; Fuks et al., 2000; Dunican et al., 2008)

1.4.6.1 Methyl-CpG Binding proteins

The first protein to specifically bind methylated CpGs, MeCP2, was identified 20 years ago (Meehan et al., 1989; Lewis et al., 1992; Meehan et al., 1992). The domain responsible for localising MeCP2 to methylated DNA was named the methyl-CpG binding domain (MBD) and was subsequently identified in several other proteins including MBD1, MBD2, MBD3 and MBD4, which are collectively known as the MBD protein family (*Fig 1.6*) (Nan et al., 1993; Cross et al., 1997; Hendrich & Bird, 1998). Of these proteins, only mammalian MBD3 lacks the capacity to bind methylated CpGs due to a substitution of a key binding residue (Saito & Ishikawa, 2002). The structure of the MBD domain complexed with methylated DNA suggested MBD proteins would recognise methyl-CpG regardless of genomic context (Ohki et al., 2001). However, it has since emerged that MeCP2 preferentially binds methyl-CpGs flanked by A/T rich sequences and that MBD1 may bind

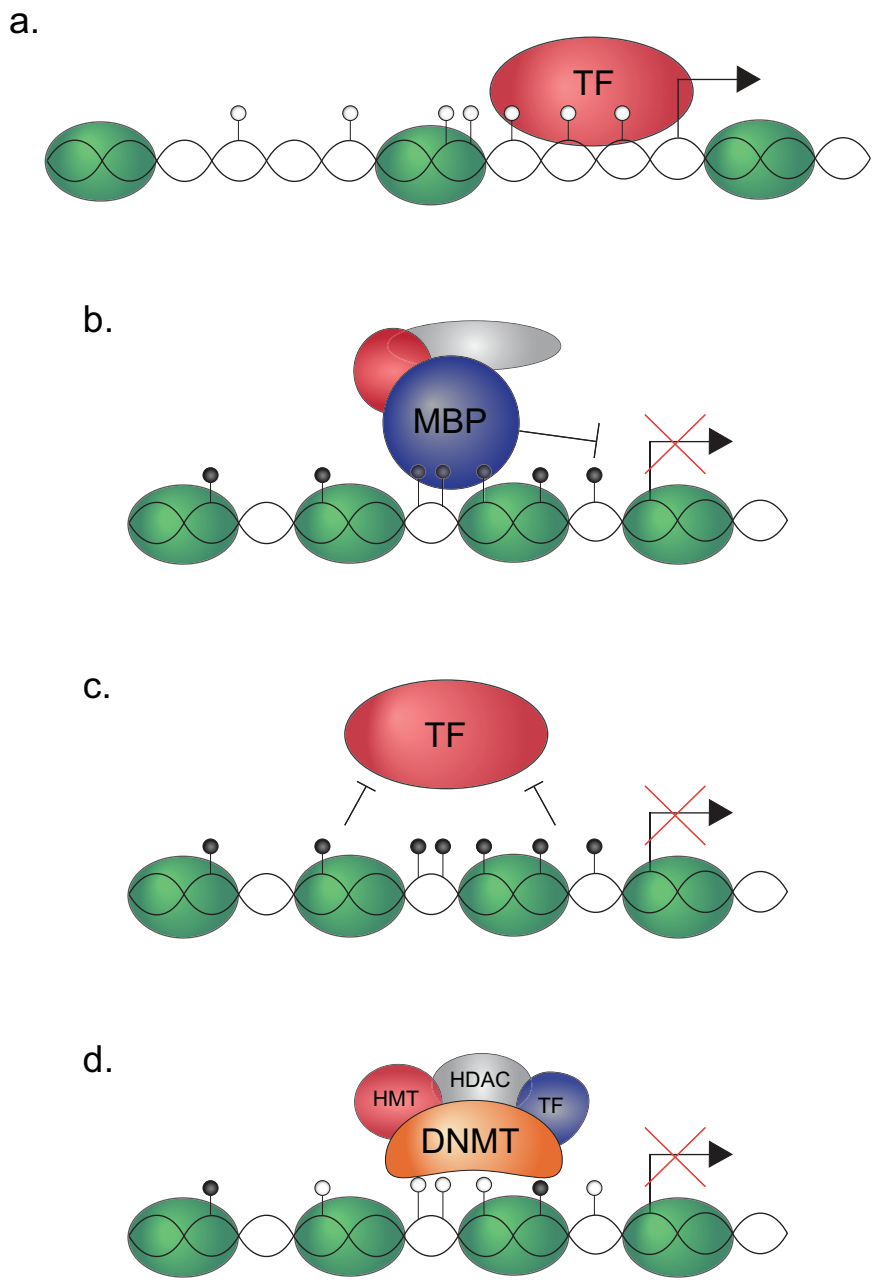


Fig 1.5. **Putative mechanisms of DNA methylation-mediated repression.** Unmethylated CpGs are shown as white circles and methylated CpGs as black circles. Green elipses represent core histones complexed with DNA. a.) Unmethylated promoters are permissive to transcription in the presence of appropriate transcription factors (TF). However DNA methylation can lead to repression and chromain condensation through one of several putative mehanisms including b.) Methyl-CpG binding proteins (MBP) can recognise methylated DNA and target co-repressor complexes to silence transcription and modify the local chromatin. c.) DNA methylation in the recognition site of some transcription factors can block them from binding DNA. This can directly inhibit transcriptional activation d.) DNA methyltrasferases (Dnmt) can recruit co-repressors such as histone deacteylases (HDAC) and histone methyltransferases (HMT) to repress transcription during *de novo* or maintenance methylation activity. This mode of repression may also occur independently of methyltransferase activity.

unmethylated CpGs (Klose et al., 2005; Jorgensen et al., 2004). MBD proteins repress transcription, at least in part, through targeting histone deacetylases, lysine methyltransferases and chromatin remodelling factors to methylated DNA which promotes an inactive chromatin structure (Nan et al., 1998, Ng et al., 2000). Each MBD protein interacts with characteristic co-repressors, although it is unclear whether all co-factors are in the same complex. For example MeCP2 is found associated with ATRX, the Sin3a/HDAC complex and H3K9 methyl-transferase activity (Jones et al., 1998; Fuks et al., 2003; Nan et al., 2007) whereas MBD1 typically associates with H3K9 methyl-transferases SETDB1 and SU39H1, and HP1 (Fujita et al., 2003; Sarraf & Stancheva, 2004).

A second group of proteins, the zinc-finger family, comprising Kaiso, Zbtb4 and Zbtb38, have also been identified as having methyl-CpG binding activity (*Fig 1.6*). These proteins recognise methylated DNA through a conserved zinc-finger domain and similarly to MBDs, target transcriptional repression to these regions (Prokhortchouk et al., 2001). The founding member, Kaiso, was initially identified in a screen for proteins that interact with the p120 catenin but was also independently recognized as binding to the *S100A4* gene in a methyl-dependent manner (Daniel & Reynolds, 1999; Prokhortchouk et al., 2001). Unlike MBDs, Kaiso was found to preferentially bind two consecutive methyl-CpGs. Kaiso interacts with the co-repressor complex NCoR via its BTB domain and may mediate transcriptional repression through recruiting this complex to methylated target genes, such as *MTA2* (Yoon et al., 2003; Collins et al., 2001). In addition to binding methylated CpGs, Kaiso binds an unmethylated TCCTGCNA sequence (KBS) as is present in the *matrilysin* promoter and promotes repression (Spring et al., 2005). However, this property is not evolutionarily conserved and in *Xenopus*, the phenotype of xKaiso depletion is rescued by zebrafish Kaiso, which binds methylated DNA but not the KBS (Ruzov et al., 2009a; Ruzov et al., 2009b). The remaining members of this family, Zbtb4 and Zbtb38, also possess a BTB domain but have additional zinc-fingers through which they bind single methyl-CpGs and repress transcription (Filion et al., 2006). As of yet, no target genes have been identified as repressed by any of the kaiso-like proteins in a normal *in vivo* developmental context (Sasai & Defossez,

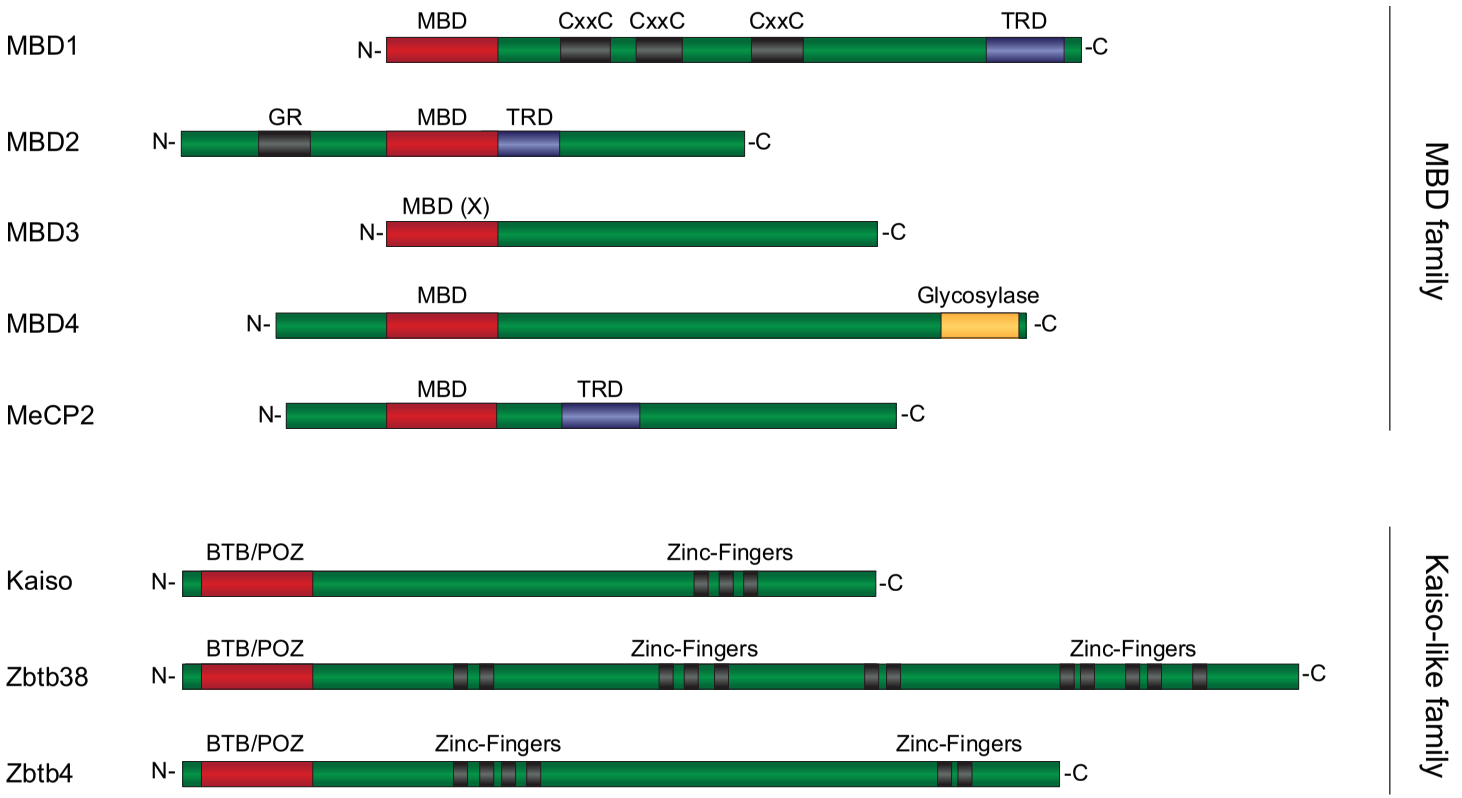


Fig 1.6. **Structure and organisation of methyl-binding proteins.** Schematic of mouse methyl-binding proteins (MBPs). The methyl-CpG binding domain (MBD) proteins bind methylated DNA via thier conserved MBD domain, with the exception of MBD3, which contains critical amino acid substitutions. MBD1 additionally contains three CxxC DNA binding domains which can target unmethylated CpGs. MBD1, MBD2 and MeCP2 contain trans-repression domains (TRD) and MBD2 contains a GR repeat motif. MBD4 contains a gylcosylase domain that contributes to its function in DNA repair. The Kaiso-like proteins are recruited to methyl-CpGs through multiple zinc-finger domains. These proteins target repression through co-repressor activity and protein interactions at their BTB/POZ domain.

2009). It is noteworthy that a third family of methyl-binding proteins, the SRA (SET and Ring associated) containing Uhrf1 and Uhrf2, have been reported to preferentially bind hemi-methylated DNA through their SRA domain. However, this family has been implicated in the deposition of DNA methylation rather than interpreting it and carrying out a biological response (Bostick et al., 2007; Sharif et al., 2007; Unoki et al., 2004).

Despite the apparent role of MBPs in repressing transcription of methylated templates *in vitro*, each MBP tested thus far is dispensable for development with mutant mice only exhibiting mild phenotypes (*Table 1.2*) (Hendrich et al., 2001; Chen et al., 2001; Guy et al., 2001; Zhao et al., 2004; Prokhortchouk et al., 2006). Additionally, in their absence only a limited number of genes are de-repressed and only to a modest extent (Barr et al., 2007; Kriaucionis et al., 2006; Sasai & Defossez, 2009). In contrast, targeted deletion of Dnmts and loss of DNA methylation leads to embryonic lethality and global gene re-activation (Li et al., 1992; Okano et al., 1999, Jackson-Grusby et al., 2001). If MBPs are mediating the effects of CpG methylation, then why do these phenotypes differ so markedly? One explanation is that there is functional redundancy between MBPs allowing them to compensate for deletion of one another. An argument against this premise is that the combined disruption of MBD2, MeCP2 and Kaiso does not significantly alter the phenotype (Martin-Cabellero et al., 2009; Guy et al 2001). However, with ever increasing numbers of methyl-binding proteins being discovered, even this triple knock out may not reveal the redundant functions of MBPs. An alternative explanation for MBP-null viable mice is that DNA methylation *per se* could be sufficient to maintain gene silencing without MBP recruitment. In this scenario, MBPs may function as an additional feedback mechanism to reinforce the repressed state but are not crucially required. It is also conceivable that MBPs do not have a global role in regulating gene expression states at all. However, work in *Xenopus* has suggested Kaiso, at least, has an important role. Here, morpholino knockdown of Kaiso has a similar phenotype to Dnmt1 knockdown, including ectopic de-repression of zygotic genes prior to the mid-blastula transition and embryonic lethality (Ruzov et al., 2004; Kim et al., 2004b; Ruzov et al., 2009a). Thus, while MBPs remain the favoured mechanism

through which repression by DNA methylation is mediated, the extent to which these proteins actually implement gene silencing remains unclear.

1.4.6.2 Direct Exclusion

A second mechanism through which DNA methylation influences transcription is direct exclusion. Here, methylated sequences inhibit regulatory proteins from binding to their cognate DNA target sequence. This mechanism is distinct from MBD recruitment in that in prevents trans-acting factors from exerting influence rather than recruiting proteins that modify chromatin structure. Such a situation is apparent with CTCF, a chromatin boundary insulator (Bell et al., 1999). Methylation prevents CTCF from binding to DNA, resulting in altered interactions between regulatory elements and promoters or modified chromatin domains. At the chicken β-globin insulator, methylation of just a single, specific CpG dinucleotide within the CTCF insulator abrogates binding (Renda et al., 2007). In mammals, this binary switch regulates the *Igf2* and *H19* imprinted genes that are expressed from parental specific alleles (*Section 1.4.5.2*) (Hark et al., 2002). In principle, all CTCF binding sites with consensus variants containing CpG dinucleotides possess the potential to regulate chromatin structure via DNA methylation. However, few examples of this kind of regulation have been reported to directly influence gene expression. A limited subset of other transcription factors have also been characterised that are blocked by DNA methylation, however, the effects on gene expression of their exclusion are unclear (Watt & Molloy, 1988; Tate & Bird, 1993; Jaenisch & Bird, 2003).

1.5 Developmental epigenetic dynamics

The cellular transition from totipotency to terminal differentiation can be defined by changes in gene expression. Thus, as echoed by Wolf Reik (2007), mammalian development is by definition, "epigenetic". That is, the orchestrated changes in gene expression that direct development largely occur without altering the DNA sequence. Molecularly, the changes in expression are coordinated by transcription factors and epigenetic systems. These processes enforce heritable transcriptional memories that stably maintain lineage commitment. The developmental acquisition of stable epigenetic marks thus results in a progressive restriction of cellular potential, 'locking' in' cell fate. In contrast, re-acquisition of developmental potential (in the zygote, iPS cells and PGCs) is accomplished by specialised epigenetic reprogramming events that reset the epigenome to a pluripotent state (*Fig 1.1*) (Hochedlinger & Plath, 2009). The epigenetic dynamics during development therefore define the phases of differentiation and reprogramming that typify the mammalian life cycle.

1.5.1 Epigenetic dynamics during embryonic development

The fertilization of an oocyte by a single sperm generates a unique totipotent cell with unparalleled cellular plasticity. To acquire this totipotent state, the genome undergoes dramatic changes in genome-wide DNA methylation patterns and histone modifications soon after fertilization. In mice, DNA methylation is actively erased from the paternal genome coincident with the removal of protamines and loading of chromatin onto histones (*Fig 1.7 upper panel*) (Mayer et al., 2000; Oswald et al., 2000). The demethylase responsible for active demethylation is unknown but may operate through a mismatch or base-excision repair mechanism due to the stable nature of the covalent bond between methyl groups and cytosine residues (Reik, 2007). In contrast, the maternal genome undergoes a slower, passive demethylation process throughout the subsequent cleavage divisions (Santos et al., 2002). The passive reduction of DNA methylation from the maternal genome is largely due to the exclusion of Dnmt1 from the nucleus at this stage (Howell et al., 2001).

However, the mechanism that protects the maternal genome from *active* demethylation has not been identified although it could be a consequence of the different chromatin organisations of the two pronuclei (Martens et al., 2005; Govin et al., 2007). Despite the global erasure of methylation after fertilization some sequences, such as allele specific ICRs and IAP elements, retain high levels of CpG methylation (Brandeis et al., 1993; Lane et al., 2003). Interestingly, Stella (also known as Dppa3) may have a role in protecting these sequences, as the absence of Stella from the zygote leads to aberrant loss of methylation from several imprinted loci and IAP elements (Nakamura et al., 2007). Functionally, the precise importance of zygotic demethylation to development is unclear, as the process is conserved between cow, mouse and pig but absent in sheep and rabbit, at least at the resolution of the 5' methyl cytosine antibody (Beaujean et al., 2004, Meehan et al., 2005).

Following implantation, the inner cell mass (ICM) but not the trophectoderm (TE), undergoes global *de novo* re-methylation (*Fig 1.7*). The differential methylation of the ICM and TE occur coincident with the first zygotic lineage restriction that distinguishes the embryonic and trophoblast lineages, respectively (Santos et al., 2002). Global levels of DNA methylation remain lower in extra-embryonic lineages compared to embryonic tissues throughout development but gene promoters may be an exception to this general depletion of methylation (Farthing et al., 2008). By the late blastocyst stage, the ICM separates into the epiblast lineage that will form the embryo, and the primitive endoderm that will form the visceral and parietal endoderm layers which contribute to extra-embryonic lineages. Post-gastrulation, *de novo* methylation and demethylation only occur at specifically targeted developmental and tissue-specific regions in somatic cells, although it is seen more widely in the development of cell lines *in vitro* (Kawai et al., 1994) and cancer cells (Jones & Laird, 1999; Keshet et al., 2006). An exception to this is germ cells, which undergo further phases of epigenetic reprogramming.

1.5.2 Epigenetics of Primordial Germ Cells

Primordial germ cells (PGCs) are the founder population of germ cells, which have the dual responsibility of differentiating into a specialised cell type (mature gametes)

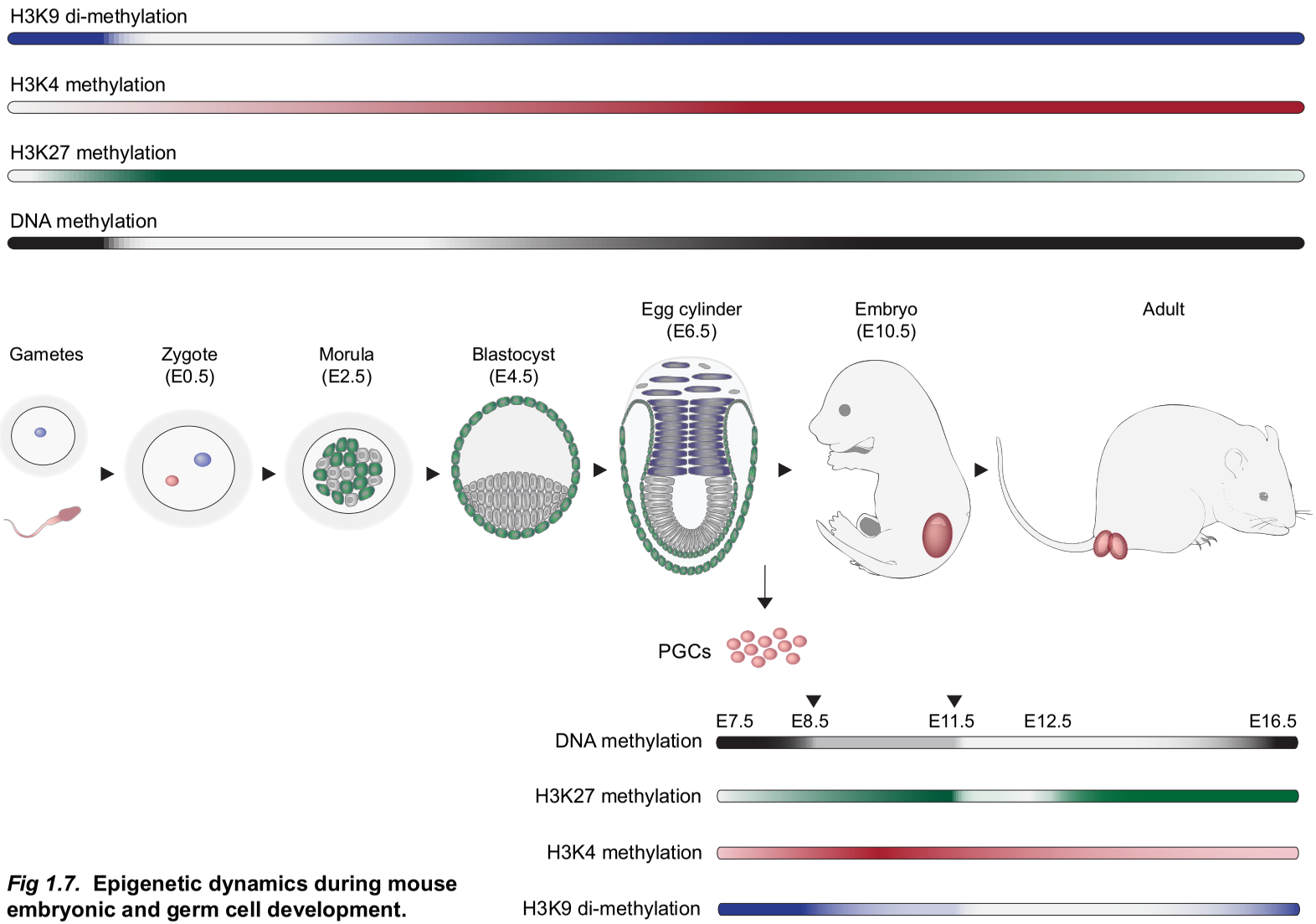


Fig 1.7. **Epigenetic dynamics during mouse embryonic and germ cell development.**

Fig 1.7. **Epigenetic dynamics during mouse embryonic and germ cell development. Top panel, embryonic epigenetics**. Very soon after fertilization DNA methylation is actively erased from the paternal pronucleus (shown as red genome/sperm) while the female pronucleus (blue genome) undergoes slower passive demethylation. H3K9me3 undergoes similar active erasure in the paternal pronucleus thus establishing an epigenetic asymmetry between the zygotic maternal and paternal genomes. In the absence of DNA methylation and H3K9 methylation, developmental gene repression is maintained by enhanced polycomb activity, particularly H3K27 methylation. *De novo* DNA methylation begins before the blastocyst stage and is largely complete by E6.5 with the genome fully re-methylated by late gastrulation (E7.5). Methylation may lock in stable silencing of pluripotency and lineage specific genes. Bottom panel, male PGC epigenetics. Primordial germ cells (PGCs) are specified at ~E7.25 in the extraembryonic mesoderm and begin migrating through the hindgut to the embryonic gonad. At ~E.8.25 they undergo an initial epigenetic reprogramming event (left arrowhead) that includes partial erasure of DNA methylation and loss of H3K9me2 concomitant with gain of H3K4- and H3K27 methylation. Shortly after PGCs reach the gonad at ~E10.5 they undergo a second wave of epigenetic reprogramming that results in complete erasure of DNA methylation and transient loss of many histone modifications including H3K9- and H3K27 methylations and histone acetylations (second arrowhead). These genomic reprogramming events are necessary for expression germline specific and pluripotency genes and the eventual acquisition of totipotency.

and retaining the capacity to form a totipotent zygote and ultimately a new organism. PGCs are specified at ~E7.25 as a cluster of cells that emerge inside the extraembryonic mesoderm at the posterior end of the primitive streak (Ginsburg et al., 1990; Saitou et al., 2002). Prior to this, precursor cells expressing *Blimp1,* which is essential for PGC specification*,* can be identified as early as E6.25 (Lawson & Hage, 1994; Ohinata et al., 2005; Vincent et al., 2005). In *Blimp1*-deficient embryos, PGC-like cells express many somatic associated Hox genes, suggesting Blimp1 is crucial for silencing of the somatic programme (Ohinata et al., 2005). Following their specification, PGCs undergo two distinct phases of epigenetic reprogramming, which promote expression of germline and pluripotency genes, and repression of a somatic cell fate.

1.5.2.1 Migratory PGCs

The first PGC reprogramming event occurs during migration through the embryonic hindgut towards the developing gonads. Here, by E8.75, nascent PGCs show a significant loss of global H3K9me2 levels and DNA methylation compared to somatic neighbours, as determined by immunostaining (*Fig 1.7 lower panel*) (Seki et al., 2005; Seki et al., 2007). Loss of these marks is accompanied by cell cycle arrest at G2 and acquisition of high H3K27me3 levels by E9.5, which may complement the erasure of DNA methylation and H3K9me2 to maintain a repressive chromatin structure (Hajkova et al., 2008; Seki et al., 2007). The relative plasticity of H3K27me3 may also be required for PGCs to eventually regain totipotency. Migratory PGCs at this stage are also enriched for the activating marks H3K4me2/3 and H3K9 acetylation relative to somatic neighbours, which may be important for expression of pluripotency factors such as Oct4, Stella and Nanog. Moreover, the enrichment of H3K4 and H3K27 methyl marks in PGCs prior to arrival at the gonad (at ~E10.5), leads to a striking resemblance to the chromatin state in ES cells (*Fig 1.7*) (Sasaki & Matsui, 2008). Currently, this reprogramming phase has not been directly associated with specific changes in gene expression and may be 'priming' the epigenetic landscape for later events. As such, the epigenetic changes that occur

in migratory PGCs may represent lineage specific (germ cell) differentiation rather than a reprogramming event *per se.*

1.5.2.2 Post- Migratory PGCs

Subsequent to entering the gonad at around E10.5, PGCs undergo another, more extensive phase of epigenetic reprogramming. This is characterised by genome wide active DNA demethylation between E11.5 and E12.5, including erasure of imprints and reactivation of the X_i chromosome (although reactivation may initiate during the migratory stage) (Hajkova et al., 2002; Monk & McLaren, 1981; Tam et al., 1994; de Napoles et al., 2007). Global levels of many repressive modifications including H3K9me3, H3K27me3 and linker histone H1 also decline at E11.5 (*Fig 1.7 lower panel*) (Hajkova et al., 2008). Intriguingly, loss of these histone marks is transient as they are regained globally by E12.5. This dynamic change in histone modifications is not just limited to repressive marks as the activating modification H3K9ac is also transiently lost at this stage. The transient global changes are accompanied by a decondensation of chromatin as revealed by enlarged nuclei. These data indicate that there is an extensive stripping of epigenetic information to completely reprogram the PGC genome at E11.5 to a pre-totipotent state. The dynamics of histone H2A.Z replacement in PGCs at this stage suggests the reprogramming occurs through global histone replacement, possibly via the histone chaperone NAP1 (Hajkova et al., 2008). Moreover, chromatin remodelling and loss of histone marks in PGCs is preceded by erasure of DNA methylation, consistent with a model whereby DNA-repair-driven demethylation directly induces histone replacement and chromatin state changes. This suggests that erasure of DNA methylation may be the initiating event for reprogramming at this stage. A minority of sequences, such as IAPs avoid complete reprogramming, however most studied regions are subject to resetting (Hajkova et al., 2002; Lane et al., 2003). One study has reported that demethylation and expression of *Dazl, Mvh* and, *Sycp3* is induced following entry into the gonad suggesting that global erasure of DNA methylation regulates the timing of activation of these germline specific genes (Maatouk et al., 2006).

In contrast to the transient loss of histone modifications, *de novo* re-methylation of PGCs does not commence until E14.5 in the male germline and not until postnatal oocyte growth in the female germline (*Fig 1.7*) (Ueda et al., 2000; Li et al., 2004; Lucifero et al., 2004). Remethylation of PGCs allows sex-specific imprints to be established at ICRs. Additionally, some genic germline methylation patterns are established, particularly at germ cell specific CGI genes, which remain hypomethylated compared to somatic cells. The imprinted methylation patterns at ICRs are catalysed by the Dnmt3a-Dnmt3L complex, with the exception of the paternal *Rasgfr1* imprint, which additionally requires Dnmt3b (Kaneda et al., 2004; Hata et al., 2002; Kato et al., 2007). The criteria by which imprinted loci are recognised and targeted in PGCs at this stage is unknown. However, the Dnmt3a-Dnmt3L complex preferentially binds CpGs spaced with 10-nucleotide intervals, which are found at many imprinted loci (Jia et al., 2007). Combined with unmethylated H3K4, which promotes Dnmt3L binding, this may provide enough specificity for targeting *de novo* activity to imprinted regions (Ooi et al., 2007). In support, a recent report noted that oocytes deficient in the H3K4me2 demethylase *KDM1b* failed to methylate several imprinted loci (Ciccone et al., 2009). The PGC-specific epigenetic patterns established in the gonad, are likely necessary for the specialised downstream functions of germ cells. Indeed promoter methylation patterns in EG cells and sperm resemble those of ES cells suggesting that while differentiating PGCs are a highly specialised cell type their genome is at least partly reprogrammed to a pluripotent state (Farthing et al., 2008).

1.5.2.3 Mature germ-cell epigenetics

Proliferation of PGCs ceases at ~E13.5 coincident with the onset of sexual differentiation. At this stage, female germ cells enter prophase I of meiosis (Tam & Snow, 1981) whereas male germ cells enter mitotic arrest by E14.5, probably as a result of signals originating from the testis cord (McLaren, 1983). In male mice, spermatogenesis resumes around 3 days post partum (3dpp) and follows a defined differentiation pathway with the first haploid spermatids present from 19dpp. Epigenetically, global chromatin remodelling occurs at this stage with the histone-

protamine exchange being the principal event. The global incorporation of basic protamines into sperm chromatin promotes DNA compaction which is both necessary for spermatozoa head size and to protect the genome from environmental insults (Sasaki & Matsui, 2008). However, despite the global protamine exchange, recent studies have noted that histones remain enriched at certain loci. Specifically, in human spermatozoa sequences with strong developmental ontologys including *Hox* genes, imprinted genes and developmental transcription factors were associated with regions of enriched histone deposition (Hammoud et al., 2009). Interestingly, histones were modified with H3K27me3 when present at loci that would be repressed in the early embryo and reciprocally H3K4me2/3 was present at future active or bivalent loci. Many developmental promoters were also DNA hypomethylated which contrasts with bulk sperm DNA which is hypermethylated. A second study using soluble and insoluble sperm chromatin fractions noted that histones were preferentially retained at promoter regions associated with CTCF binding sites (Arpanahi et al., 2009). Taken together these studies demonstrate that sperm DNA contains at least two chromatin packaging domains and that epigenetic marking in sperm is extensive. These modifications may have a role in directing the global reprogramming following fertilization or in excluding certain loci from reprogramming to promote heritable gene expression states.

1.5.3 Epigenetic Systems during development

During development different subsets of genes are expressed to define the cellular lineage and potential. To regulate the dynamic expression of gene subsets, epigenetic processes utilize both short-term flexible systems and stable long-term heritable mechanisms (Reik, 2007). For example, because development is a progressive restriction of potential, genes required for early pluripotency are conceptually obsolete after somatic differentiation and therefore can be permanently silenced. In contrast, developmental genes are not required during early development but are necessary at later stages and thus need to be transiently repressed but ‘poised’ for expression without the requirement for epigenetic reprogramming. To achieve these distinct outcomes, different epigenetic systems operate which mediate either flexible

repression and activation or heritable silencing during development. These systems regulate different gene subsets depending on the developmental requirement and thus, epigenetic marks accumulate differentially according to cell-type and lineage. Reprogramming events enable the erasure of these marks and thereby allow the cycle of restriction of cell plasticity by these permanent and flexible epigenetic systems to repeat.

1.5.3.1 Flexible epigenetic regulation

In mammals, the flexible repression of developmental associated genes, such as the *Hox, Pou* and *Gata* families, in pluripotent and multipotent cell-types is achieved through the Polycomb group (PcG) proteins (Lee et al., 2006). These proteins form two complexes, PRC2 and PRC1, which deposit H3K27me and H2A119ub1 to cooperatively mediate gene repression at targeted developmental loci and therefore maintain pluripotency. The importance of PcG proteins was demonstrated through ES cells deficient in the *Eed* or *Suz12* components of PRC2, and therefore which lacked H3K27me3. Here, many developmental genes were de-repressed and the mutant ES cells were liable to spontaneously differentiate (Boyer et al., 2006; Lee et al., 2006). This is consistent with a model whereby PcG complexes maintain ES cell pluripotency and plasticity during embryonic development through transiently repressing developmental genes. The intriguing observation that that many PcG target genes were marked by extended regions of H3K4me3 as well as H3K27me3 - so called bivalent domains - led to the concept that these genes were poised for activation (Bernstein et al., 2006). In this model, the repressive H3K27me3 is dominant over H3K4me3 and therefore bivalent genes are transiently silenced until lineage specific differentiation induces depletion or removal of H3K27me3. Upon lineage specific loss of the repressive mark, genes transition from a 'poised' to an 'active' state due to the primed H3K4me3 modification (Bernstein et al., 2006; Mikkelsen et al., 2007; Pan et al., 2007). Whether H3K27me3 is removed by down regulation of PRC components or an active demethylase at individual loci is unknown, but crucially, the mechanism allows repressive marks to be flexibly and quickly removed in response to downstream cues (Mohn et al., 2008; Barski et al.,

2007; Zhao et al., 2008). Unlike *Drosophila*, a general polycomb response element (PRE) has not been identified in mammals and therefore it is unclear precisely how PcG proteins target specific developmental genes. One mouse sequence (PRE-*kr*) has recently been shown to recruit PcG complexes to loci containing the motif in an orientation specific manner and therefore may be a mammalian specific PRE (Sing et al., 2009). However, as the sequence is not present at all target loci, alternative targeting mechanisms must also be in operation (Schwartz & Pirrota, 2008).

1.5.3.2 Stable epigenetic regulation

In contrast to developmental genes that are transiently repressed until required, pluripotency genes are stably epigenetically inactivated as cellular differentiation proceeds because in principle, they are only required in the germline. These genes are characteristically marked by repressive histone modifications, DNA methylation and replicate during late S-phase (Lande-Diner & Cedar, 2005). The classical example is *Pou5f1* (encoding Oct3/4*)* (*see Section 1.4.4*), which acquires histone modifications and DNA methylation in a defined epigenetic cascade that results in permanent silencing of the locus (Feldman et al., 2006). Several distinct mechanisms cooperate to reinforce long-term stable silencing at pluripotency loci to prevent reactivation. The collaborative effort to maintain permanent silencing is probably due to the hazardous effects of ectopic reactivation of pluripotency genes in differentiated cell types such as a potential predisposition to cancer (Hochedlinger et al., 2005). Permanent epigenetic silencing is also utilized during development to preserve the fixation of cell fate. For example, heritable methylation of the *Elf5* gene leads to stable silencing which prevents embryonic lineages from entering a trophoblast lineage fate (Ng et al., 2008; Hemberger et al., 2009). Thus, in general, distinct epigenetic systems modulate the activity of genes based on the developmental requirement for flexibility or stability of gene expression and/or lineage commitment.

1.6 Pluripotent stem cells

1.6.1 Pluripotent cell types

1.6.1.1 ES cells

In 1981, two groups simultaneously isolated and cultured embryonic stem (ES) cells derived from the inner cell mass (ICM) of mouse blastocysts (Martin, 1981; Evans & Kaufman, 1981). Mouse ES (mES) cells have both the capacity to self-renew and to differentiate into lineages that represent every tissue found in an adult mouse. *In vivo* this pluripotent state exists only transiently, in the developing blastocyst. However *in vitro*, mES-cell pluripotency can be sustained indefinitely through continual exposure to defined factors (Chambers & Smith, 2004). Indeed, even following extensive culture, mES cells reintroduced into developing blastocysts retain their capacity to contribute to all adult tissues in chimeric animals, although earlier passage cells have a higher propensity to be transmitted through the germline. They additionally contribute to allantois, yolk-sac mesoderm and amnion (Bradely et al., 1984). However, in accordance with their epiblast origin and unlike human ES cells (hES), mES cells contribute relatively insignificantly to the primitive endoderm and never to the trophectoderm (Beddington & Robertson, 1989). This partial lineage restriction is why mES cells are regarded as pluripotent, whereas early zygotic cells, which can contribute to all extra-embryonic lineages, are totipotent. Despite this, the capacity of mES cells to colonise the germ cell lineage in chimeric animals, makes them highly effective tools for manipulating the mouse genome and ultimately for generating mutant mice and cell lines (Downing & Battey Jr, 2004).

Original lines of mES cells were maintained on cultures of irradiated or mitotically blocked ‘feeder’ fibroblasts. The observation that the critical pluripotency factor secreted by fibroblasts was leukaemia inhibitory factor (LIF), has led to the generation of feeder-free cell lines cultured in the presence of LIF and serum (Williams et al., 1988; Smith et al., 1988). In turn, this paradigm was enhanced upon discovery that serum in the media can be supplanted by bone morphogenetic proteins (BMPs), which function in collaboration with LIF to maintain mES cell self-renewal

(Ying et al., 2003a). Most recently the generation of 3i media, which inhibits FGF receptor tyrosine kinases, the ERK cascade and glycogen synthase kinase-3 (GSK3), has enabled continuous culture of mES cells capable of self renewal (Ying et al., 2008). Because 3i media inhibits differentiation inducing FGF signals rather than promoting pluripotency *per se*, the development of this media demonstrated self-renewal is an intrinsic property of the mES-cell state rather than a consequence of pluripotency promoting signals.

Due to the relatively recent development and cost of 3i media most studies still utilize LIF for work with mES cells. Mechanistically, LIF functions through the gp130 receptor to trigger STAT3 in a JAK kinase modulated system (Niwa et al., 1998). Withdrawal of LIF from cultured mES cells promotes spontaneous differentiation along multiple lineage paths. However, this can be used in combination with refined protocols to induce differentiation into a variety of cell-types representing all three germ layers, but rarely trophoblast cells (Yasunaga et al., 2005; Nishikawa et al., 1998; Ng et al., 2008; Chambers & Smith, 2004). A key differentiation method entails allowing mES cells to aggregate in suspension leading to structures called embryoid bodies (EB). This process mimics the process of embryogenesis and therefore differentiation into EBs is thought to occur on a comparable schedule to *in vivo* development, but without the structure of the developing embryo. Interestingly, a recent report has demonstrated that embryoid bodies establish an anteroposterior polarity and form a primitive streak-like region supporting the premise that EB formation closely resembles *in vivo* development (ten Berge et al., 2008). Most alternative methods differentiate mES cells during monolayer culture or on matrices, which lessens the parallels with embryonic development but allows enhanced access and manipulation of cultures. Here, addition of varying concentrations of growth factors (Okabe et al., 1996) or retinoic acid (Rohwedel et al., 1999) influences differentiation toward specific lineages. While these general methods result in a heterogeneous assortment of differentiated cells, innovative techniques have also been employed to attain a homogeneously differentiated populations (Rolletschek et al., 2001; Ying et al., 2003b).

1.6.1.2 Other pluripotent cells

In addition to the generation of mES cells from blastocyst stage embryos, other phases of mouse development allow the derivation of alternative pluripotent cell lines *in vitro*. Pluripotent embryonic carcinoma (EC) cells can be established from epiblast teratomas but contribute poorly to chimeras and rapidly lose plasticity in culture due to a propensity to be aneuploid (Hogan et al., 1994; Evans, 1972; Roassant & McBurney, 1982). More regularly, pluripotent embryonic germ (EG) are generated from the primordial germ layer between E8.5 and E12.5, with the efficiency of derivation being significantly higher from earlier stages (Matsui et al., 1992; Durcova-Hills et al., 2001). EG cells have comparable developmental potency to ES cells and express the SSEA1 antigen, tissue non-specific alkaline phosphotase and *Oct3*/4, all of which are markers of pluripotency (Matsui et al., 1992). However, EG cells derived from post-migratory stage PGCs lack the correct complement of imprints leading to skeletal abnormalities in chimeras (Tada et al., 1998; Shamblott et al., 1998; Spivakov & Fisher, 2007). Recently, spermatagonial stem cells (SSCs) from adult males have also been derived and cultured (Kubuta & Brinster, 2006). Most recently, epiblast stem cells (EpiSC) have been derived from E5.5 to E6.5 rat and mouse epiblast tissue (Brons et al., 2007). EpiSC are able to differentiate to all three germ layers and form teratomas in immunodeficient mice but contribute very poorly to chimeras, possibly due to a partially limited capacity to develop into early lineages. However, recent evidence has suggested that epiSCs can de-differentiate to ES-like cells (rESCs) through extended culture with LIF suggesting that they are not developmentally limited (Bao et al., 2009). Interestingly, both the expression profile and maintenance culture requirements suggest mouse epiSC are more comparable to human ES cells than to mES cells (Tesar et al., 2007; Brons et al., 2007).

1.6.2 Maintenance of Pluripotency

1.6.2.1 Pluripotent Transcription Factor Networks

Attempts to characterise the mediators of self-renewal have examined both the epigenetic landscape and the transcriptional profile of ES cells. Analysis of the

genetic basis of pluripotency has identified a subset of genes exclusively expressed in mES cells (Ivanova et al., 2002; Imamura et al., 2006). Further work revealed at least three factors – Oct3/4, Sox2 and Nanog – form a core network of 'master regulators' that are important for pluripotency in mES cells (Boyer et al., 2005; Nichols et al.,1998; Chambers et al., 2003; Avilion et al., 2003). Genome-wide ChIP on chip analysis of these 'regulators' demonstrated cooperative binding by Oct4, Nanog and Sox2 at a sizeable number of developmental loci (Loh et al., 2006; Sharov et al., 2008). Interestingly, Oct4 and Nanog are associated with gene promoters irrespective of expression state, suggesting these key regulators can bind active as well as repressed targets. However, genes shown to be repressed were principally developmental loci involved in lineage commitment, whereas active genes bound by Oct4/Nanog were associated with an uncommitted state and expressed in cell-types able to self-renew (Loh et al., 2006; Boyer et al., 2005). Thus, this pluripotency network activates genes that promote self-renewal while silencing differentiation associated loci. Recently other factors such as Sall4 and Klf4 have been identified that extend this network of core pluripotency regulators (Zhang et al., 2006; Yang et al., 2008). Additional reports have indicated that multiple core factors (greater than 4) cooperatively activate key pluripotency loci whereas genes with few core factors bound are generally repressed (Kim et al., 2008; Chen et al., 2008). It will therefore be important to delineate the precise role of each pluripotency factor and their interactions with one another in the future, as it has recently been shown that Nanog is not strictly required for maintenance of pluripotency at all in cultured mES cells (Chambers et al., 2007)

Interestingly, many factors that comprise the core pluripotency network regulate their own expression either directly or indirectly by modulating the expression of epigenetic modifiers (Wu et al., 2006; Chew et al., 2005; Rodda et al., 2005; Loh et al., 2007). This creates a self-reinforcing loop of pluripotency but also suggests that disturbances to this equilibrium could cause a binary shift leading to repression of pluripotency factors and differentiation. For example, JMJD2C maintains *Nanog* expression by reversing H3K9me2 at the *Nanog* promoter but expression of JMJD2C itself is dependent on Oct3/4 levels which in turn are modulated by the G9a-EHMT1

complex directing H3K9me2 to the *Oct3/4* promoter (Loh et al., 2007; Feldman et al., 2006; Hattori et al., 2007). Therefore, the dynamic balance between epigenetic modifiers and the core pluripotency factors maintains the pluripotent state. This highly-tuned system could explain the susceptibility of ES cells to spontaneously differentiate *in vitro* and the transient nature of the pluripotent state during development where stochastic events constantly affect the transcriptional equilibrium (Hemberger et al., 2009).

1.6.2.2 Epigenetic basis of pluripotency

The epigenetic landscape of pluripotent cells functionally cooperates with the core pluripotency transcription factors to maintain self-renewal. Principally, the H3K27me3 mark catalysed by polycomb group proteins maintains the transient repression of differentiation associated genes in pluripotent cells therefore promoting plasticity in culture and during embryonic development (Boyer et al., 2006). The responsive nature of this system is, at least partly, due to the epigenetic flexibility of the bivalent marks found predominately in ES cells. Interestingly, ~30% of genes are not marked by either bivalent modification but are found to be DNA methylated (Fouse et al., 2008). DNA methylation at these loci may represent a parallel system to polycomb that heritably stabilises gene expression and self-renewal in ES cells. However, as *Dnmt1*-null ES cells can proliferate and self-renew, the precise role of DNA methylation at these loci with respect to maintaining pluripotency is unknown.

1.6.2.3 Epigenetic Reprogramming of somatic cells

Mammalian development is a unidirectional process during which epigenetic systems and transcription factors retain the memory of lineage specification. However, seminal experiments in amphibians and mammals showed the genomes of differentiated cells were able to generate viable cloned animals (Gurdon, 1962; Wilmut et al., 1997; Eggan et al., 2004). These studies indicated that the restriction of lineage potential can be manipulated and reversed suggesting epigenetic mechanisms rather than permanent genetic changes mediate developmental restriction (Hochedlinger & Jaenisch, 2006). More recent work has demonstrated that

differentiated nuclei can be dominantly reprogrammed to express markers of earlier embryonic stages by fusion with ES cells (Tada et al., 2001). These observations led to the astounding observation that temporally restricted ectopic expression of the quartet *Oct4, Sox2, Klf4* and *c-Myc* can reprogram somatic fibroblasts to induced pluripotent (iPS) ES-like cells (Takahashi & Yamanaka, 2006; Takahashi et al., 2007). Completely reprogrammed iPS cells exhibit transcriptional profiles and histone modification patterns very similar to ES cells. Moreover, there is erasure of DNA methylation at key pluripotency genes and reactivation of the inactive X chromosome in female cells (Maherali et al., 2007; Mikkelsen et al., 2008; Okita et al., 2007). Importantly, iPS cells can generate viable chimeric embryos and contribute to the germline demonstrating their functional pluripotency (Okita et al., 2007; Wernig et al., 2007). Recent studies have demonstrated iPS cells can be derived from many cell types (Aoi et al., 2008; Wernig et al., 2008; Eminli et al., 2008) and with alternative combinations of factors, such as *Esrrb* instead of *Klf4* or by excluding *c-myc* entirely (Feng et al., 2009; Nakagawa et al., 2008). The efficiency of iPS cell derivation can also be increased by manipulating the epigenetic landscape with inhibitors of DNA methylation and histone lysine methylation (Mikkelsen et al., 2008; Shi et al., 2008). By transiently reducing DNA methylation it is hypothesised that stable silencing of pluripotency genes is weakened allowing more efficient reversal of lineage commitment. These data illustrate the central role of epigenetic processes and transcription factors in defining cellular memory. Further understanding of the epigenetic landscapes that dictate expression states and cellular memory will vastly broaden our medical and conceptual capacity. Indeed, epigenetically reprogrammed iPS cells have already led to potential patient-specific therapies for sickle cell anaemia (Hanna et al., 2007) and Parkinson's disease (Wernig et al., 2008).

1.7 Thesis Objectives

Multiple epigenetic systems cooperate with trans-acting effectors to precisely regulate cellular patterns of gene expression. Determining how these mechanisms operate and interact is important for our understanding of normal developmental processes, transcriptional memory and ultimately human disease. This thesis aims to investigate the role of one major epigenetic modification, DNA methylation. DNA methylation has been well characterised as the primary mediator of mono-allelic gene expression, transposon silencing and maintenance of X-chromosome inactivation. However, the function of DNA methylation with respect to epigenetic regulation of single-copy genes is contentious. Central to the debate is the lack of evidence that differential DNA methylation patterns are the cause, rather than the consequence, of transcriptional activity. This continues to sustain the argument that, at best, promoter methylation is recruited by other epigenetic systems to 'lock in' a long-term silenced state and therefore is not a mechanism for regulating expression *per se* but a secondary maintenance system. I set out here to determine whether DNA methylation is the primary system for regulating the expression of any single-copy genes and thus the maintenance of transcriptional memory at these loci.

The aims of this thesis are thus:

1. To identify candidate genes potentially regulated primarily by DNA methylation.

2. To evaluate the functional role of promoter methylation at candidate genes in a developmental context.

3. To investigate the mechanism through which promoter methylation silences candidate genes.

Chapter 2

Materials and Methods

2.1 Bacterial Manipulation

2.1.1 Transformation of bacterial cells

Chemically competent bacterial DH5α *Escherichia coli* (*E. coli*) (Invitrogen) were incubated with 10-100ng of DNA (not greater than 5% bacterial volume) on ice for 30 minutes. The transformation mixture was heat shocked at 42°C for 45 seconds and cooled rapidly on ice for a further 2 minutes. Room temperature S.O.C rescue media was added to a final volume of 500µl and the mixture was incubated at 37°C with shaking for 1 hour. Transformed cells were then aseptically spread onto L-agar plates with the appropriate antibiotic added (Ampicillin at 50mg/ml, Kanamycin at 50mg/ml, Chloramphenicol at 13mg/ml), inverted and incubated at 37°C overnight. For blue/white selection 80µl X-gal (25mg/ml) was pre-streaked onto the agar plates.

2.1.2 Bacterial culture and isolation of plasmid DNA

For the preparation of plasmid DNA, transformed DH5α cells were incubated at 37°C overnight with shaking (225rpm) in 5ml liquid Luria-Bertani (LB) culture medium containing the appropriate antibiotic. For maxi-preparations, 1ml of this starter culture was inoculated into 150ml of antibiotic containing LB and incubated at 37°C for a further 16 hours. Plasmid DNA was extracted with either the Nucleospin Plasmid kit (Machery-Nagal) for mini-preparations or the GenElute HP *Endotoxin-Free* Maxiprep Kit (Sigma) for maxi-preparations according to the manufacturer's instructions.

2.2 Mammalian Cell Culture

2.2.1 Cell culture

All cells were incubated in a humidified atmosphere at 37ºC with 5% CO_2. Cell lines were cultured in optimised media according to their requirements as shown in *Table 2.1*. ES cells were grown for 2-3 days in the presence of LIF and 'fed' with fresh media every day and three hours before passaging. To split ES cells, the culture media was aspirated and cell colonies were washed once in pre-warmed PBS. A small volume of trypsin-EDTA (10% v/v) was added and cells were incubated for 3 minutes at 37°C. ES cells were dislodged by gentle agitation and 10ml of pre-warmed media was added to inactivate the trypsin and create a single cell suspension. The cell suspension was counted and seeded on either pre-gelatinised (0.1% in PBS) or mitomycin C (10 µg/ml) treated MEF monolayer culture flasks at a concentration of 1.5 x 10^5 cells/ml, with 15ml added to a T75cm^2 flask. All other cell lines were passaged every 2-3 days at a ratio of 1 in 5 to 1 in 10 as ES cells but without gelatinised or feeder layer culture flasks and with media replaced only during passaging (*Table 2.1*).

Cell type	Genotype	Strain	Culture Media	Passaged
ES	WT	J1	GMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 0.1mM ß-Mecaptoethanol: 1000U/ml LIF; 1000U/ml Pen; 650µg/ml Strep	1 x 10^5/ml gelatin
ES	WT	E14'	DMEM; 10% FCS; 1xNEAA; 2mM L-Glutamine; 0.1mM ß-Mecaptoethanol: 1000U/ml LIF; 1000U/ml Pen; 650µg/ml Strep	1 x 10^5/ml gelatin
ES	WT	E14	GMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 0.1mM ß-Mecaptoethanol: 1000U/ml LIF; 1000U/ml Pen; 650µg/ml Strep	1 x 10^5/ml gelatin
ES	*Dnmt3a*$^{-/-}$	J1	GMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 0.1mM ß-Mecaptoethanol: 1000U/ml LIF; 1000U/ml Pen; 650µg/ml Strep	1 x 10^5/ml gelatin
ES	*Dnmt3b*$^{-/-}$	J1	GMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 0.1mM ß-Mecaptoethanol: 1000U/ml LIF; 1000U/ml Pen; 650µg/ml Strep	1 x 10^5/ml gelatin
ES	*Dnmt3a*$^{-/-}$; *Dnmt3b*$^{-/-}$	J1	GMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 0.1mM ß-Mecaptoethanol: 1000U/ml LIF; 1000U/ml Pen; 650µg/ml Strep	1 x 10^5/ml gelatin

ES	*Eed*$^{-/-}$		DMEM; 10% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 0.1mM ß-Mecaptoethanol: 1000U/ml LIF; 1000U/ml Pen; 650µg/ml Strep	1 x 10^5 /ml feeders
ES	*Hdac1*$^{-/-}$	E14'	DMEM; 10% FCS; 1xNEAA; 2mM L-Glutamine; 0.1mM ß-Mecaptoethanol: 1000U/ml LIF; 1000U/ml Pen; 650µg/ml Strep	1 x 10^5 /ml gelatin
pMEF	*WT*		DMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 6
MEF	*p53*$^{-/-}$		DMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 8
MEF	*Dnmt*$^{-/-}$; *p53*$^{-/-}$		DMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 5
MEF	*Lsh*$^{-/-}$		DMEM; 15% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 5
Fibroblast	WT		DMEM; 10% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 8
Fibroblast	*MBD2*$^{-/-}$		DMEM; 10% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 8
Fibroblast	*Kaiso*$^{-/-}$		DMEM; 10% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 8
Fibroblast	*Kaiso*$^{-/-}$; *MBD2*$^{-/-}$; *MeCP2*$^{-/-}$		DMEM; 10% FCS; 1xNEAA; 2mM L-Glutamine; 1mM Na Pyruvate; 1000U/ml Pen; 650µg/ml Strep	1 in 8
Neuro2a	WT		DMEM; 10% FCS; 1xNEAA0; 1000U/ml Pen; 650µg/ml Strep	1 in 8
293T	WT		DMEM; 10% FCS; 1000U/ml Pen; 650µg/ml Strep	1 in 10

***Table 2.1*. Cell culture conditions and genotypes.**

2.2.2 Freezing & thawing

Cell suspensions were frozen in 1ml aliquots of cell culture media supplemented with 10% (v/v) dimethyl sulfoxide (DMSO) in liquid nitrogen (N_2). To facilitate a slow freezing process, cell aliquots were initially frozen at –20ºC for 2hrs, then transferred to –70ºC for 24hrs before being deposited in liquid N_2. Cells were retrieved from liquid N_2 by thawing in a 37ºC water bath, followed by dropwise dilution of the cell suspension in 10ml pre-warmed (37ºC) culture medium. The cells were collected by centrifugation, to remove the DMSO, and seeded into 25cm^2 culture flasks. Flasks for ES cell culture were pre-coated with sterile 0.1% gelatin in PBS or mitomycin C treated primary MEFs (*Table 2.1)*.

2.2.3 Cell counting

Cells were harvested into a single cell suspension in PBS for counting. An aliquot of 10ul was pipetted into a Neubauer haemocytometer (0.1mm depth, 1/400mm^2) and the cell number in each of four 1mm^2 areas was counted, averaged and multiplied by 10^4. This number was multiplied appropriately for the dilution in PBS to obtain the number of cells per ml in the suspension.

2.2.4 DNA transfection

Cells were seeded at least 24hrs prior to transfection in culture media lacking antibiotics and grown to a confluence of 70-80%. Plasmids were transfected into cells with lipofectamine 2000 (Invitrogen) in accordance with manufacturer's instructions. Briefly, plasmid DNA and lipofectamine were incubated separately in an appropriate volume of Opti-mem (Gibco) for 5 minutes and then pooled with gentle mixing. The combined solution was incubated for a further 25 minutes before being added directly to cells. Lipofectamine 2000 was used at a ratio of 2.5µl per 1µg of plasmid DNA except where stated.

2.2.5 Soft agar growth assay

A basement layer of 0.5% agar (autoclaved) made with DMEM was set into 60mm culture plates. 1 x 10^5 cells in the appropriate culture medium and 0.3% agar were plated in suspension on the basement layer and incubated in a humidified atmosphere at 37°C. After 12 days, colonies were stained with 0.1% (w/v) crystal violet (in 20% ethanol) and counted.

2.2.6 Poly(HEMA) anchorage-independent growth assay

Poly(2-hydroxyethyl methacrylate) (poly(HEMA)) prevents anchorage-dependent growth and therefore only transformed cells are able to proliferate on poly(HEMA) coated surfaces. To prepare plates, 50ul poly(HEMA) (5mg/ml in ethanol) (Sigma) was pipetted into 96-well tissue culture plates and allowed to air-dry at 37°C for

24hrs. For control, half the wells contained ethanol only. Coated plates were sterilized with ultraviolet (UV) immediately prior to use. For the measurement of anchorage-independent cell proliferation, 2.5 x 10^3 cells ($p53^{-/-}$, $dnmt1^{-/-}$ $p53^{-/-}$, pMEF, or 293T) were seeded per well, in triplicate, for each time point in a final volume of 150ul. The appropriate culture media for each cell type was used but phenol red was omitted as it interferes with absorbance readings. At each time point for growth to be measured, 15ul 3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) (5mg/ml in PBS) was added and the cells were further incubated at 37°C for 4hrs. The resulting MTT formazon was solubalised by adding SDS to a final concentration of 8% (w/v) and incubated at 37°C for a further 16 hrs. Because yellow coloured MTT is reduced to purple formazon only in living (viable) cells, measurement of absorbance wavelengths between 520-600nm gives a quantitative value for cell proliferation rates. The absorbance was measured at 570nm and a reference wavelength of 690nm using a microplate reader **(**FluoSTAR optima F). The reference wavelength value was subtracted from the 570nm absorbance and averaged across the triplicate wells to give a growth ratio at each time-point +/- poly(HEMA).

2.2.7 5-aza dC treatment and cellular recovery

When cells reached a confluence of 40-50%, the culture media was replaced with media containing 1uM or 10uM 5-aza dC (10mM stock in 50% acetic acid solution) (Sigma) or control vehicle. Because 5-aza dC is highly unstable in solution, fresh media was added every 24hrs during treatment periods. Following treatment, cells were either harvested for downstream experiments or further cultured and passaged through a recovery period for up to 14 days in the absence of 5-aza dC.

2.2.8 Differentiation of ES cells into embryoid bodies

Confluent ES cells were trypsinised and diluted in differentiation media (appropriate cell media but lacking LIF) to a concentration of 3 x 10^4 cells per ml. Using a multi-channel pipette, 120 drops of 20ul (~600 cells/drop) were applied to the lid of a 100mm^2 square bacterial (non-adhesive) culture dish. The lid was inverted over a

pool of 10ml PBS and incubated at 37°C. Approximately 4 dishes were required per time point for RNA/DNA analysis. After 2 days growth in hanging drops cells had aggregated into embryoid bodies (EB). EBs were collected by washing and soft centrifugation, and placed in suspension culture in bacterial dishes for a further five days. A half-media change was made every day and EB's were transferred to a fresh bacterial culture dish every second day. At day 7 the EB's were transferred to a mammalian cell culture dish and allowed to adhere. Correct differentiation of EB's was scored morphologically by the presence of rhythmically beating colonies, which typically appeared after 9-10 days and represent cardiomyocyte lineage differentiation.

2.2.9 Retinoic acid differentiation of ES cells

All ES cells were seeded at a concentration of 1 x 10^5 per ml in gelatinised T75cm^2 flasks (20,000cells/cm^2). After 12 hours growth in the presence of LIF, cells were washed twice with PBS and fresh differentiation media supplemented with retinoic acid (RA) (10^{-6}M) (Sigma) was added. RA culture media was prepared in low light conditions and changed every 24 hours during the treatment period.

2.2.10 in vitro embryonic gonad culture

The gonads of male and female CD1 mice were dissected from day E13.5 embryo's (n = 30 x male, 32 x female) (*Dr Ian Adams*). Gonads were incubated on 8cm^3 tissue culture (T/C) 1% agar blocks made up with PBS in a 35mm culture dish at 37°C. The culture dish was filled with PGC media (DMEM, 10% FCS, 2mM L-Glutamine, 1mM Na Pyruvate; 0.1mM ß-Mecaptoethanol, 1000U/ml Pen; 650µg/ml Strep, 1 x Fungizone) and adjusted to just above the level of the upper plane of the T/C blocks. Fresh media +/- 1uM 5-aza dC was added every day. Gonads were fixed for *in situ* hybridisation after 48hrs (E15.5) or 72hrs (E16.5) of *in vitro* growth.

2.2.11 Generation of stable cell lines

The plasmid to be integrated was linearised at a unique site in a non-essential region for downstream applications (*NotI* for pZX-hDNMT1^{C1226Y}) and purified by phenol/chloroform extraction. Linearised plasmids were transfected into cells with either Lipofectamine 2000 or by microporation. Transfected cells were grown for 48hrs and then split at a ratio of 1 in 20 with the appropriate selection agent (Zeocin at 300μg/ml, Puromycin at 2.5μg/ml) added to select for stable integration of the plasmid. The optimal concentration of selection agent for selection of low copy number integrations was determined by a prior kill curve. Transfected cells were maintained under selection for 12 days with fresh media added every 2-3 days. After 12 days, colonies which had expanded from a single cell clone were scored for the presence of absence of GFP which is present in the pZX vector. GFP positive, selection resistant colonies were picked, trypsinised and seeded into 96-well plates to allow clonal expansion. Once cells had reached sufficient numbers (~1x10^5 cells) they were scored for stable integration by RT-PCR for the gene of interest (hDNMT1 or hDNMT1^{C1226Y}).

2.2.12 Generation of primary Mouse Embryonic Fibroblasts (MEFs)

E13.5 mouse embryos were dissected into PBS and the embryonic internal organs were removed from the abdominal cavity. The embryos were dissociated by aspirating into a 10-ml syringe through a fine 16-G needle and expelling the contents. Trypsin/EDTA solution was added to 1 x concentration and the cells were incubated at 37°C for 5 minutes followed by vigorous pippeting. An equal volume of primary MEF (pMEF) culture media was added and the remaining large tissue pieces were allowed to settle in a 50ml conical tube. The supernatant was centrifuged and re-suspended in 15ml fresh media per embryo. The cell suspension was plated into T75cm^2 flasks using approximately one embryo per flask and grown until confluent.

2.3 DNA preparation and manipulation

2.3.1 Gel electrophoresis

DNA samples were resolved by horizontal electrophoresis through Hi-pure low EEO agarose gels ranging between 0.8-2% supplemented with 0.4µg/ml ethidium bromide. For expected DNA fragments <800bp, gels were made in conjunction with 0.5 x TBE, for DNA fragments >800bp gels were made with 1 x TAE. Samples were loaded with the appropriate volume of 6x loading buffer (0.4% orange G, 0.03% bromophenol blue, 0.03% xylene cyanol FF, 15% Ficoll® 400, 10mM Tris-HCl (pH 7.5) and 50mM EDTA (pH 8.0) and the resulting DNA fragment sizes were estimated with either GeneRuler 1kb ladder (Fermentas), 0.07-1.2kb ladder (Roche) or 100bp DNA marker (Promega)

2.3.2 DNA fragment extraction from agarose gels

DNA fragments were purified from gels using the Nucleospin II extract kit (Machery-Nagal) according to the manufactures instructions. Briefly, transverse gel slices containing the DNA fragment(s) of interest were cut from the gel with a scalpel blade and dissolved in 2 gel slice volumes of buffer QT at 50ºC for 10 minutes. The sample was centrifuged at 11,000g through the silica-gel column membrane included in the kit. This column retains DNA with high affinity at the elevated salt concentrations used here. Bound DNA was washed twice to remove impurities and dried by centrifugation for 1 minute at 11,000g. DNA was eluted with 15-50µl of 10mM Tris depending on the required final concentration.

2.3.3 Phenol/Chloroform and ethanol precipitation

An equal volume of 25:24:1 phenol:chloroform:isoamyl alcohol (50% buffered phenol, 48% chloroform [v/v]. 0.5% 3-methyl-1-butanol [v/v] pH >7.8) was added to at least 100µl of nucleic acid sample, pulse vortexed and centrifuged at 16,000g for 5 minutes in a benchtop microcentrifuge. Aqueous and organic components divided into phases and the top aqueous layer was taken to the next stage of the process. To

remove residual phenol, an equal volume of chloroform was added to the aqueous sample, pulse vortexed and centrifuged at 16,000g for 5 minutes. The top aqueous phase was recovered and DNA was precipitated by addition of $1/10^{th}$ volume of 3M sodium acetate and 2 volumes of ice-cold ethanol. For limiting quantities of DNA (<1µg) 1µl of glycoblue (15mg/ml, Ambion) was added to the precipitation. To recover the DNA, the sample was centrifuged at 16,000g for 20 minutes at 4ºC. The pellet was washed with 70% ethanol to remove residual salt, re-pelleted by centrifugation and left to air-dry at room temperature. The washed pellet was re-suspended in the desired volume of MilliQ ultrapure dH_2O.

2.3.4 Nucleic acid quantification

The concentration of nucleic acids was measured spectrophotometrically with a nanodrop ND-1000. 1µl of sample was loaded and the absorbance of the sample measured at a wavelength of 260nm (A_{260}). An A_{260} value of 50 equal's 1ug/ml of double stranded DNA or 40ug/ml single stranded RNA. A_{260}/A_{280} and A_{260}/A_{230} ratios were recorded to determine the purity of the sample.

2.3.5 pGEM T Easy cloning

Because *Taq* polymerases have a 3' terminal deoxynucleotidyl transferase activity, they frequently add deoxyadenosine to 3' ends of PCR products. The linearised pGEM-T easy vector (Promega) exploits this through having 3'deoxythymidine overhangs, which promote efficient ligation of PCR products. Purified PCR product was mixed with 50ng pGEM-T easy at an approximately 3:1 molar ratio and incubated in 1x ligation buffer (Promega) and 3 Weiss Units of T4 DNA Ligase for 1 hour at room temperature or overnight at 4°C. The ligation reaction was used to transform chemically competent DH5α *E.Coli.*

2.3.6 Sequencing

DNA was sequenced with the Big Dye™ terminator v3.1 sequencing kit according to the manufacturer's instructions, with exception that 2µl of Big Dye was used in a

10µl reaction instead of 4µl. Products were ethanol precipitated and run on an ABI Prism 3730 genetic analyser. Sequence data was analysed with Bioedit Sequence Alignment Editor v9.0.9.

2.3.7 in vitro methylation

Plasmids were incubated at 37ºC for 2hrs with the appropriate bacterial methylase and buffer (NEB) supplemented with 160µM S-Adenosyl methionine (SAM) or 32µM SAM for HhaI methylase. The breakdown product of SAM, S-adenosyl-L-homocysteine (SAH), inhibits the methyltransferase reaction. To shift the equilibrium towards driving the reaction, fresh SAM was added after 2hrs and the sample incubated for a further 1hr at 37ºC. Products were purified by phenol/chloroform extraction (*Section 2.2.1*) and the *in vitro* methylation was confirmed by digestion with the methyl-sensitive enzyme *HpaII* and its methyl-insensitive isoschizomer *MspI*.

2.3.8 Southern blotting

Samples were run on a large 1% agarose gel at 4°C and post stained with ethidium bromide. The gel was washed twice for 15 minutes in denaturing SS1 solution (0.5M NaOH, 0.5M NaCl) and then twice in neutralising SS2 solution (0.02M NaOH 1M AmmAc). The gel was pre-soaked in 20 x SSC and the DNA was transferred to a positively charged nylon membrane (Pal) overnight by capillary action. The membrane and gel were placed between Whatman paper saturated in 20 x SSC with the lower surface dipped into a reservoir of 20 x SCC. To draw the buffer upwards and transfer the DNA to the membrane, a wad (~15cm thick) of dry towels and weight was placed on the upper Whatman. Following confirmation of successful transfer, the membrane was baked at 80°C for at least 3 hours to immobilize DNA. The membrane was probed with [α ^{32}P] dCTP (50uCi, GE Healthcare) labelled pMR150, which hybridises to mouse minor satellite. Membranes were exposed on phosphor screens and visualised with a FLA-5100 phosphorimager (reytek) and Advanced Image Data Analyser (AIDA) version 3.44.035.

2.4 RNA preparation and analysis

2.4.1 Isolation and purification of RNA

When cell numbers were not limiting (>1 x 10^4), Trizol reagent (Invitrogen) was used to purify RNA. Trizol was either added directly to cells in culture (after removing media) (1ml per 25cm^2) or used to re-suspend trypsinised embryonic tissue. Trizol is a monophasic solution of phenol and guanidine isothiocyanate which disrupts cells and cell components while maintaining the integrity of RNA. After 2 minutes at room temperature to disrupt nucleoprotein complexes 1/10^{th} volume of bromochloropropane (BCP) was added and the solution mixed vigorously for 15 seconds. Separation of the aqueous and organic phases was achieved by centrifugation at 12,000g for 15 minutes at 4°C in a microfuge. The RNA containing upper, aqueous phase was transferred to a fresh RNase free eppendorf and precipitated with an equal volume of isopropanol (and 1ul glycoblue (Ambion) for limiting quantities of RNA) for 10 minutes at room temperature. RNA was pelleted by centrifugation at 12,000g for 12 minutes and washed with 75% ethanol before re-suspension in RNase-free ultra-pure MilliQ dH_2O and long-term storage at –80°C. For RNA isolation from small cell numbers (<1 x 10^4), such as flow sorted E10.5 germ cells, the RNeasy mini kit (Qiagen) was used according to the manufacturer's protocol

2.4.2 Preparation of cDNA

For the generation of complimentary DNA (cDNA) all RNA samples were DNase treated with Turbo DNA-free DNase (Ambion) according to the manufacturer's instructions. Where possible, 5ug RNA was used in each DNase reaction. First strand cDNA synthesis was carried out with superscript II or superscript III reverse transcriptase (RT) (Invitrogen) and was primed with random hexamers. For a 20ul reaction, ~1.5ug DNased RNA, 150ng random primers and 1ul dNTPs (10uM) were heated to 65°C for 5 minutes. This reaction was then supplemented with first strand buffer (to 1x), 1ul DTT (0.1M), 1ul RNase inhibitor (Promega) and 200U Superscript III RT and incubated at 25°C for 5 minutes, 50°C for 1 hour and 70°C

for 15 minutes to heat in activate the RT. In parallel a reaction of pooled RNA's without RT was performed as a control for contaminating DNA or template aerosols. Reverse transcribed cDNA's were diluted to 50ul with ultra-pure MilliQ dH_2O and stored at –20°C short term (<2 months) or -70°C for long term storage.

2.4.3 RT-PCR

To determine RNA transcription levels from genes, RT-polymerase chain reaction (RT-PCR) was used. RT-PCR was performed with 1.75mM $MgCl_2$, 0.25uM each primer, 200nM dNTPs, 1 x PCR buffer (20 mM Tris-HCl (pH 8.4), 50 mM KCl.) and 1U platinum *Taq* (Invitrogen) using the first strand cDNA as a template. Platinum *Taq* is complexed with a proprietary antibody until heat-activated and therefore doesn't promote primer-dimer formation, which can strongly influence PCR kinetics. RT-PCR parameters were specific for primer sets (*Appendix 1*) but the majority were designed to perform optimally under these cycling conditions:

95°C for 2 minutes
94°C for 15 seconds ┐
56°C for 20 seconds │ 18-35 cycles
72°C for 25 seconds ┘
72°C for 10 minutes

2.4.4 microRNA detection and quantification

Testis specific microRNA's (miRNA) were analysed using the High-Specificity miRNA qRT-PCR Detection Kit (Stratagene) according to the manufacturer's instructions. Briefly, 500ng total RNA was polyadenylated at the 3' end with poly A polymerase (PAP), to elongate short miRNAs. As a control, parallel reactions were carried out without PAP. Polyadenylated or control RNA was then used as a template for first strand cDNA synthesis with the reaction primed by an RT adapter primer that anneals to the 3′ poly-A tail. The RT adaptor primer contains additional bases that create a universal sequence tag at the 5′ end of each cDNA. For quantitative

detection of miRNA targets, real-time qPCR (EvaGreen) (Stratagene) was used with a universal reverse primer that binds to the universal tag at the 5' end of cDNAs. The specificity of the qPCR reaction was provided by the miRNA-specific forward primer. This primer was designed to be identical in sequence and length to the miRNA of interest. However, to ensure mispriming to similar miRNA's was avoided, a bioinformatic analysis of all known mouse miRNA's was completed to identify testis-specific miRNA's that differ from all other miRNA's by at least 2 nucleotides in the 3' region and 3 nucleotides in total. A further analysis with Stratagene's online miRNA primer design guide was also undertaken to ensure specificity. For miRNA microarray analysis, total RNA extracted from each cell line was quantitated and analysed for integrity. Labelling and hybridisation to the custom miRNA expression array was performed at Leicester University (*Dr Emma Wade*).

2.4.5 Whole mount In situ *hybridisation*

Whole mount *in situ* hybridization was carried out with Tex19 3'UTR cRNA probes labeled with digoxigenin For probe generation, full-length Tex19 3'-UTR was transcribed from the T7 or SP6 promoters of linearised plasmid (pGemTEasy-Tex193'UTR) to generate sense and antisense transcripts respectively. The transcription reaction was carried out with 500ng linearised plasmid, T7 or SP6 polymerase, appropriate transcription buffer and the DIG RNA labeling kit (Roche) according to manufacturer's instructions for 2 hours at 37°C. RNAs were purified by phenol/chloroform extraction and ethanol precipitation and aliquoted with DEPC treated MilliQ ultrapure dH_2O.

For *in situ* hybridization freshly dissected E13.5 gonads or *in vitro* cultured E15.5 and E16.5 gonads were fixed in 4% paraformaldehyde (PFA) (in PBS) at 4°C overnight. Fixed gonads were washed twice with PBS then dehydrated through 50% methanol (in PBS 0.1% tween (PBSt)) and 100% methanol washes and stored at -20°C for less than 1 month. Gonads were rehydrated through sequential washes of 75%, 50%, 25% methanol (in PBSt) and twice with PBSt. Rehydrated embryo's were incubated with 10ug/ml Proteinase K (PBSt) for 14 minutes, washed, and post-

fixed with 4% PFA / 01% glutaraldehyde. Gonads were incubated with 1:1 PBSt/hybridization mix and then hybridization mix (50% Formamide, 1.3x SSC, 5mM EDTA, 50ug/ml Yeast RNA, 0.2% Tween-20, 0.5% CHAPS, 100ug/ml Heparin in DEPC dH_2O) for 5 minutes. The gonad samples were then incubated with fresh hybridization mix for 1 hour at 68C before pre-warmed hybridization mix with 1ug/ml DIG-labeled cRNA probe was added and incubated with rocking at 68C overnight. Probe bound gonads were washed 4 x 30mins with prewarmed (68°C) hybridization mix and 4 x 30 minutes with MABT (100mM maleic acid, 150mM NaCl, 0.15 Tween-20 in DEPC in dH_2O, pH7.5).

Samples were incubated with MABT / 2% Boehringer blocking reagent (BBR) for 1 hour, then 1 hour with MABT / BBR / 20% heat treated sheep serum. Following this a 1/2000 dilution of AP anti-DIG antibody (Boehringer) was added to gonads with fresh MABT / BBR / 20% serum and incubated overnight at 4°C. Gonads were then rinsed and washed 3 x 1 hour with MABT and 2 x 15 minutes with NTMT (100mM NaCl, 100mM Tris pH 9.5, 50mM $MgCl_2$, 1% Tween-20 in DEPC dH_2O). To develop colour gonads were incubated with 1.5ml NTMT with 4.5ul/ml NBT (75mg/ml in DMFO) and 3.5ul/ml BCIP (X-phosphate; 50mg/ml in 70% DMFO) for 1 hour to 5 hours in the dark. Once colour developed gonads were rinsed with PBSt and re-fixed in 4% PFA at 4°C overnight.

2.5 Protein preparation and analysis

2.5.1 Cell extracts

Whole cell protein extracts were prepared by washing cells in PBS and adding an appropriate volume of Lammelli buffer (2% SDS, 10% glycerol, 5% 2-mercaptoethanol, 0.004% bromphenol blue and 60mM Tris HCl, pH 6.9). The solution was sonicated briefly to shear chromatin and boiled for 5 minutes at 95°C to denature proteins. Alternatively for cleaner protein extracts, washed cells were incubated for 15 minutes on ice in RSB-150t (0.1% tween) and spun at 4,500g for 8 minutes to (remove histones and cell component debris). The supernatant was taken and glycerol added to a 10% final concentration, before storage at –20°C

2.5.2 SDS-PAGE

Cell protein extracts were resolved by denaturing sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Briefly, a separating gel with 10-12% acylamide (v/v), 375mM Tris-HCl (pH8.8), 0.1% SDS (w/v), 0.075% Ammonium persulphate (APS) (w/v), 0.15% N, N, N', N'-Tetramethylethylene-diamine (TEMED) (v/v) was prepared and allowed to set with a isopropanol overlay. Acrylamide was from 30% (29:1 acrylamide/bis-acrylamide (w/v)) stock (Severn Biotech). The isopropanol was rinsed away and a 4% acrylamide stacking gel with 125mM Tris-HCl (pH 6.8), 0.1% SDS (w/v), 0.15% APS (w/v), 0.3% TEMED (v/v) was set on top with loading wells. Samples and pre-stained protein ladder (Fermentas) were loaded and the gel run in SDS-PAGE running buffer (25mM Tris-HCl, 200mM Glycine, 0.1% SDS (w/v), pH 8.8) at 160V for ~1 hour.

2.5.3 Western blotting

Following protein separation by SDS-PAGE the stacking gel was removed and proteins were transferred to a nitrocellulose membrane (Amersham Hybond ECL) with a genie blotter (Idea Scientific) in transfer buffer (25mM Tris, 200mM Glycine, 20% methanol) according to the manufacturer's instructions. For proteins >150kDa, transfer buffer was reduced to 10% methanol and 0.1% SDS was added. Following transfer for ~45 minutes at 30amps membranes were blocked in 4% non-fat dry milk proteins (in PBS) (Marvel) for 1 hour at room temperature. Primary antibodies (anti-Tex19 1:100 (*Dr Ian Adams*), anti-Dnmt1 1:1500 (Abcam), anti-Tubulin 1:2000, anti-T7 1:10,000 (Novagen)) were added at the appropriate dilution in fresh 1% milk proteins (in PBS / 0.2% tween (PBSt)) and incubated overnight with agitation at 4°C. Membranes were washed 3 x 5 minutes in PBSt before addition of the appropriate secondary antibody conjugated to horse-radish peroxidase in 1% milk proteins/PBSt for 1 hour at room temperature with constant agitation. Non-bound antibody was washed 3 x 15 minutes with PBSt and membranes were developed using chemilumuminescent detection (SuperSignal Western Pico Reagent, Pierce) for 5 minutes. Signals were exposed on ECL Hyperfilm (Amersham-Pharmacia).

2.6 Experimental procedures

2.6.1 Real-Time quantitative PCR (qPCR)

Quantitative PCR analysis of cDNA or DNA fragments generated from ChIP was performed with Brilliant qPCR mastermix II (Stratagene) in a CFX96 qPCR thermocycler (Biorad). Reactions of 25ul were composed of 12.5ul mastermix, 1-2ul sample, 0.25uM each primer in MilliQ ultrapure dH_2O. For every primer set, standards were created through 10-fold serial dilutions of input (ChIP) or a high expressing sample (qRT-PCR). Standards and samples were performed in duplicate according to the following cycling parameters:

95°C for 10 minutes
94°C for 15 seconds ┐
56C for 15 seconds │ 18-35 cycles
72°C for 20 seconds ┘
60°C-95°C melt curve analysis

After each 72°C extension, the fluorescence was measured and used to calculate the cycle number at which each standard and sample crossed a fluorescence threshold (C_t). As the C_t of a reaction is determined linearly by the amount of template present at the beginning, relative quantification can be achieved by normalising to a reference gene (such as *Gapdh*) and a standard curve. The standard curve was constructed by plotting the log of the dilution factor (log^{10}) of standards against the C_t value obtained during amplification of each dilution. This allowed two outcomes. Firstly, the equation of the linear regression line, along with Pearson's correlation coefficient (r), could be used to evaluate whether the qPCR assay was optimised. For a primer set to be considered optimised they should have produced an $r > |-0.99|$, and an amplification efficiency between 90-105%. Secondly, the efficiency of both the target and reference primer set, as determined by correlation and linear regression, allowed the use of the stringent Pfaffl equation for determination of

normalised relative gene expression (Fleige & Pfaffl, 2006) This equation can be expressed as such:

$$\text{Ratio} = \frac{(E_{\text{target}})^{\Delta Ct,\ \text{target (control – sample)}}}{(E_{\text{ref}})^{\Delta Ct,\ \text{ref (control – sample)}}}$$

The Pfaffl method, unlike other methods, takes into account that the target (gene of interest) does not have the same amplification efficiency as the reference (normalisation gene) and therefore gives more accurate and reproducible results. Relative gene expression was determined based on the Pfaffl equation using CFX96 manager and *Excel 97*. To ensure specific priming, a melt curve analysis was performed at the end of each qPCR to determine whether one or several products had been generated. Only reactions in which one product, of the expected melting temperature, was generated were used for further analysis.

2.6.2 Plasmid constructs

In general, plasmid constructs were cloned via directional sticky end digestion and ligation using the rapid ligation kit (Roche). hDnmt1 was cloned into pZX in frame using the *XbaI* and *MluI* sites of pZX-Luc. Because hDNMT1 contains an internal *XbaI* site, I used *Spe1* to digest DNMT1, which creates a complimentary overhang to *Xba1,* but eradicates the overall restriction site when cloned. An additional T7 tag and kozak sequence was incorporated into the forward primer sequence. The *Xenopus Kaiso* zinc-finger domain and *GCNF* were cloned into pVP16 in frame using the *EcoR1* and *XbaI* sites. Germline specific promoters were cloned into pGL3-basic using *MluI* and *BglII* sites.

2.6.3 Isolation of mouse germ cells by fluorescent activated cell sorting (FACS)

Matings of heterozygous mice carrying an *Oct3/4* promoter cassette driving a GFP transgene, which is a specific marker for germ cells in post-blastocyst embryos, were set up. The hindgut of E9.5 embryos and the gonads of E10.5 and morphologically sexed E13.5 embryos were dissected in PBS. Samples were trypsinised for 3 minutes

and dissociated into a single cell suspension by pippeting and filtering through a 40um membrane. Cells were re-suspended in a 5% FCS/ PBS solution. Flow cytometric cell sorting was performed using a BD FACSAriaII SORP (Becton Dickinson). The 488nm laser was used for measuring forward scatter, side scatter and GFP fluorescence (525/50nm bandpass filter). BD FACSDiva software (Becton Dickinson, Version 6.1.2) was used for instrument control and data analysis. A clear bi-modal distribution was observed with GFP expressing germ cells fluorescing $>10^3$ orders of magnitude higher than testis somatic cells. GFP positive germ cells and GFP negative somatic cells were sorted into aliquots of 400 cells and at E13.5, additional aliquots of 4,000 cells. Aliquoted cells were gently pelleted, snap frozen and stored at –80°C. At E9.5 there were ~500 germ cells per embryo, at E10.5 ~1000 germ cells per embryo and at E13.5 ~10,000 germ cells per embryo.

2.6.4 Luciferase-reporter assays

Plasmids were transfected into cells in a 24 well plate as described. All wells received 30ng thymidine kinase driven renilla luciferase (Tk-Rn) and 250ng of firefly luciferase driven by a germline-specific promoter cassette. For other firefly luciferase plasmids see appendix. If cells were co-transfected with effector plasmids there were either control wells with empty plasmid or, where a dose curve was used, each well was balanced to have the same total amount of plasmid DNA. After 24 hours cells were harvested and assayed for luciferase expression with the Dual Luciferase Assay kit (Promega) according to the manufacturer's instructions. Fluorescence readings were made with a Lumat LB 9507 tube luminometer (Berthold Technologies). To account for variations in cell viability and transfection efficiency firefly luciferase was normalised to Tk-Rn.

2.6.5 Native chromatin immunoprecipitation (n-ChIP)

A total of 5×10^7 cells were harvested, washed in ice cold PBS and re-suspended in NBA/B buffer (85mM NaCl, 5.5% Sucrose, 10 mM TrisHCl pH 7.5, 0.2 mM EDTA, 0.2 mM PMSF, 1 mM DTT, 0.05% NP40, 1X Protease Inhibitors) for 3 minutes on ice to disrupt cell membranes. Nuclei were centrifuged at 2,500rpm for 4

minutes at 4°C and re-suspended in NBR (85 mM NaCl, 5.5% Sucrose, 10 mM TrisHCl pH 7.5, 3 mM $MgCl_2$, 1.5 mM $CaCl_2$, 0.2 mM PMSF, 1 mM DTT) and re-centrifuged at 2,500rpm for 4 minutes at 4°C. Chromatin concentration was determined after a 5 minute DNase I digestion and measurement of the A_{260} with a spectrophotometer and diluted to 1ug/ul in 500ul aliquots. Chromatin aliquots were treated with 1ul RNase A/T1 cocktail (Ambion) for 5 minutes and then digested with micrococcal nuclease (MNase) at an optimal predetermined concentration (14U for p53-/- (P) and *Dnmt*-/- *Trp53* -/- (DP) cells) for 10 minutes at room temperature. At this concentration penta-nucleosomes were just visible by EtBr staining. The reaction was stopped by adding an equal volume of RBB buffer (215 mM NaCl, 10 mM TrisHCl pH 8, 20 mM EDTA, 5.5 % Sucrose, 2 % TritonX 100, 0.2 mM PMSF, 1 mM DTT, 2X Protease Inhibitors) and chromatin was released from nuclei overnight on ice.

The following day the sample was centrifuged at 5,000rpm for 5 minutes and the supernatant, containing the released chromatin, was transferred to a new tube. $1/10^{th}$ of the sample was kept at –20°C as the input. Prewashed and blocked protein G beads (Amersham) in a 1:1 slurry with blocking solution (0.5% BSA in PBS) were added to the chromatin and rotated for 2 hours at 4°C and the pelleted to pre-clear chromatin of non-specific interactions with beads. In parallel, 60ul of slurry was suspended in 200ul block solution with 2-10ug antibody and rotated for 2 hours at 4°C. 1ml of precleared chromatin (50-150ug) was added to the antibody bound beads and the immunoprecipitation (IP) carried out for 3hrs at 4°C. Following this, pre-optimised washes were carried out appropriate to the antibody used. Washed, immunoprecipitated material was eluted from the beads twice with 75ul elution buffer (0.1 M $NaHCO_3$, 1 % SDS) at 30°C with vortexing. The eluted sample and the input were treated with 40ug proteinase K for 1 hour at 55°C and all samples were purified with the QiaQuick kit (Qiagen) according to the manufacturer's instructions. The relative enrichment of immunoprecipitated samples was calculated by real-time qPCR using the $1/10^{th}$ input to construct a standard curve. Enrichment is expressed as a percentage of total input.

2.6.6 Crosslinking chromatin immunoprecipitation (x-ChIP)

Cells were washed twice in ice cold PBS and crosslinked with 1% PFA (in PBS) for 12 minutes at room temperature with gentle agitation. The crosslinking reaction was stopped by addition of glycine to a final concentration of 125mM. Fixed cells were washed twice with PBS and NCP buffer (200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 10 mM Hepes pH=6.5, 0.25% Triton X-100) and re-suspended in 500ul SDS lysis buffer (50 mM Tris-HCl pH=8.1, 10 mM EDTA, 0.5% NP40, 1% SDS). Lysates were sonicated with a Biorup tor NG (Diagenode) to shear DNA to ~200bp-1500bp for 3.5 minutes with 30 second on/off pulses on high as pre-determined by optimization. Sonicated samples were centrifuged at 10,000g for 10 minutes to remove cellular degree and $1/10^{th}$ inputs were removed. 225ul of IP buffer (20 mM Tris-HCl pH=8.1, 150 mM NaCl, 2 mM EDTA, 0.1% Triton X-100 (v/v), Protease inhibitors) was added to 125ul of chromatin suspension and precleared with 50ul 1:1 protein G bead (Amersham) slurry and 5ug single-stranded salmon sperm DNA (ssssDNA) (Sigma) for 2 hours with rotating at 4°C. Preclearded supernatant was supplemented with 100ul IP buffer and 1-5ug appropriate antibody and immunoprecipitated overnight at 4°C with rotation. The following day, 50ul protein G slurry and 2ug ssssDNA was added to each IP and incubated for a further 2-3 hours. Following this, pre-optimised washes were carried out appropriate to the antibody used. Washed, immunoprecipitated material was eluted from the beads twice with 75ul elution buffer (0.1 M $NaHCO_3$, 1 % SDS (w/v)) at 30°C with vortexing and crosslinks were reversed by incubation at 67°C overnight. Samples were purified and quantitated as N-ChIP (*Section 2.6.5*).

2.6.7 5' Rapid amplification of cDNA ends (5' RACE)

The transcription start site(s) (TSS) of *Tex19* were determined with the 5' Rapid Amplification of cDNA Ends (5'RACE) kit v2.0 (Invitrogen) according to the manufacturer's instructions. Briefly, first strand cDNA was synthesised from total RNA with a Tex19 specific upstream reverse primer (5'-TGACTCTGACAAGT ATTCC-3'). A homoploymeric cytosine tail was then added to the 3' end using terminal deoxynucleotidyl transferase (TdT). The 5' region of the Tex19 transcript

could then be PCR amplified using a nested Tex19 primer (5'-CTCTTGCCAGTC TCCCATCTC –3') and a deoxyinosine-containing anchor primer which anneals to the poly-cytosine tract and contains a 5' anchor tag. Nested PCR with a second nested Tex19 primer (5'-TGGTACAGCCATGCCTCATAG-3') and an abridged anchor primer gives a specific product (of expected size 350bp) that contains the 5' TSS of *Tex19* in each sample. This PCR product was cloned into pGEM-T easy and sequenced with Sp6 polymerase.

2.6.8 Bisulphite sequencing analysis

The methylation status of genomic CpGs was assayed by bisulphite sequencing with the EZ DNA Methylation Gold kit (Zymo Research) following the included protocol. This method distinguishes methylated from unmethylated CpG's as unmethylated cytosines are converted to uracil in the presence of bisulphite whereas methylated CpG's are protected. DNA sequencing can then identify unmethylated (thymine) and methylated (cytosine) CpGs, respectively. Where possible, 400ng genomic DNA was bisulphite treated with CT conversion reagent under the following conditions; 98C for 10minutes, 65°C for 2.5 hours, 4°C 1-12 hours and desulphinated through a silica column. Purified DNA was PCR amplified with bisulphite primers specific to the region of interest (*Appendix 1*) for 30 cycles. Where necessary, nested PCR using $1/50^{th}$ of the first PCR reaction was carried out for a further 20-35 cycles. PCR products of the expected size were gel purified, cloned into pGEM-T-Easy and transformed into DH5α *E.coli*. Colonies with an insert (white) where mini-prepped and sequenced with Sp6 polymerase. Bisulphite sequences were analysed for methylation status using Bioedit Sequence Alignment Editor v9.0.9 using published reference DNA sequences (Ensembl).

For bisulphite analysis of limiting numbers ($<1 \times 10^4$) of FACS sorted germ or testis somatic cells, an alternative method was used. Pelleted cells were directly lysed by addition of 20ul BS Prep buffer (20ug/ul Proteinase K, 1% SDS) at 56ºC for 1 hour. This solution was then added directly to CT conversion reagent and the protocol as above was followed.

2.6.9 Promoter-specific in vitro methylation

Plasmids containing the promoter of interest were *in vitro* methylated as described (*Section 2.3.7*). Following purification, the methylated promoter fragment control unmethylated fragments were digested out of the plasmid (pGL3-basic) with two unique site restriction endonucleases (*KpnI* and *HindIII*). The promoter fragments were purified by gel extraction and ligated back into unmethylated backbone vector (pGL3-basic) cut with *KpnI* and *HindII* with the rapid ligation kit (Roche). To identify re-insertions, plasmids were linearised at a non-functional site (SalI) and resolved on an agarose gel. Plasmids containing a methylated or control unmethylated promoter fragment were identified by size and by co-running a linearised starting plasmid. Plasmids of the appropriate size were gel extracted and purified by phenol/chloroform and ethanol precipitation. Approximately 15-20ng of plasmid containing a specifically methylated promoter region or control was obtained through each run. Once equalised, 10ng was transfected into each 24 well plate for luciferase reporter assays.

2.6.10 Immunofluoresence

Cells were placed on poly-lysine coated slides in a humidified atmosphere for 20 minutes or grown on ethanol-sterilized cover slips. For immunofluoresence, cells were fixed with 4% PFA for 10 minutes and washed twice with PBS before being permeabilised with 0.1% Triton-X (in PBS) for 10 minutes. Samples were blocked with 5% Donkey serum (in PBSt-Azide-BSA) for 1 hour and washed with PBS prior to addition of primary antibody in blocking solution for 1 hour at room temperature Antibodies were: rabbit anti-Oct4 (2µg/ml) (abcam - ab18976), rabbit anti-Nanog (3ug/ml) (*Dr Ian Chambers*), rabbit anti-Tex19 (5ug/ml) (*Dr Ian Adams*). Cells were washed three times in PBS and incubated with the appropriate secondary antibody (1:2000) (Alexa Fluor 488 or 594, Invitrogen) for 30 minutes. Following three final washes with PBS cells were stained with DAPI and mounted in Vectashield**.** Cells were visualised using a Zeiss Axioplan II fluorescence microscope with Plan-neofluar objectives. Image capture and analysis were performed using in-house scripts written for IPLab Spectrum (Scanalytics Corp, Fairfax, VA).

2.6.11 Expression microarray

Total RNA was extracted from each cell line at two time points six days apart for biological replicates. The integrity of RNA samples was confirmed quantitatively by on-chip electrophoresis using the Bioanalyzer 2100 (Agilent). Biotinylated cRNA for hybridisation to Mouse Ref-8 v2 or Mouse Ref-6 v2 Expression BeadChips (Illumina) was generated using the TotalPrep RNA Amplification kit (Illumina) according to the manufactures instructions. Briefly, 250ng total RNA was used to synthesise double stranded DNA (dsDNA). Purified dsDNA was used as a template for *in vitro* transcription of biotin labelled cRNA for 6 hours at 37°C. Biotinylated cRNA was purified through filter cartridges and re-analysed using the Bioanalyzer 2100 to confirm integrity. cRNA was adjusted to 150ng/ul and 5ul of each sample was hybridized to BeadChip expression arrays (performed by the Wellcome Trust Clinical Research Facility, Edinburgh). Each biological replicate was synthesised in duplicate as technical replicates so that for each sample there were two biological replicates each of which had two technical replicates (n = 4 per sample).

2.7 Bioinformatics and statistics

2.7.1 Microarray analysis

The raw data generated from microarray analysis was analysed with Beadstudio Gene Expression Module v3.4 (Illumina) using an average normalization with background removal. For differential expression analysis a t-test error model with multiple testing corrections using Benjamini and Hochberg false discovery rate was applied. The relatively high number of replicates used enabled high confidence multiple corrected t-test significance values. However in effect, this produced significant t-test differential values for genes with low fold-changes (e.g. >1.2 fold for high expressing genes).

As a primary filter to determine genes significantly differentially expressed between samples I imposed a >6-fold expression change threshold. To determine fold-changes I initially normalized and ranked transcripts according to the detection *p*-

value, which is calculated based on signals of negative controls. The detection *p*-value is calculated as *1-R/N*, where *R* is the rank of the gene signal relative to negative controls and *N* is the number of negative controls. The detection *p*-value (set here as $p < 0.05$) can be interpreted as the probability of seeing a certain signal level without specific probe-target hybridisation. To generate fold-changes it was necessary to transform normalised and ranked data to account for negative expression signals. Thus, the expression signal (*E)* at which transcripts were considered statistically detected ($p < 0.05$) was determined and all expression signals in the control sample (wild-type) >6-fold below *E* were normalised to a value equal to *E/*6. Expression fold-changes between samples were ranked and filtered according to detection ($p < 0.05$) of the transcript in at least one sample. 98.2% of transcripts that recorded a >6-fold differential expression change were significant (t-test $p < 0.05$) and 92.3% were highly significant (t-test $p < 0.01$).

2.7.2 Gene ontology (GO) and bioinformatic analysis

Functional annotations were performed using the program Database for Annotation, Visualization, and Integrated Discovery (DAVID) 2007 (Dennis et al. 2003). All analyses presented in this thesis used the same DAVID parameters. These parameters were Gene Ontology (GO) Biological Process, level 1 and 3, normalised to Illumina BeadChip v2 Ref-6 or Ref-8 backgrounds. The EASE *p*-value threshold was set at 0.05, with a minimum number of 5 genes in the ontology category. Benjamini and Hochberg multiple testing corrections were determined but not applied as each gene can contribute to multiple ontologies and therefore does not produce a binary outcome. For tissue specificity analysis, DAVID v6 was used with a *p*-value threshold set at 0.05. Venn diagrams were generated with GeneVenn (http://www.bioinformatics.org/gvenn/index.htm).

2.7.3 RT-PCR primer design

RT-PCR primers were designed using complimentary DNA sequences obtained from the ensembl genome browser. Primers were designed using Primer3 software (http://frodo.wi.mit.edu/primer3/) using the rodent and simple mis-priming library to

avoid generation of primers in degenerate or repetitive regions. Where possible primers were designed to overlap at least one exon-exon junction (intron-spanning) with a 3' terminal GC clamp. Generally primer PCR products were between 100 and 350bp. For quantitative RT-PCR some primer sets were retrieved from primerbank (http://pga.mgh.harvard.edu/primerbank/) (Spandidos et al., 2009), which contains sets of experimentally validated primers for 27,681 mouse genes. Primer sets retrieved from here were additionally validated through Primer3 and confirmed to be intron- spanning. All primer sets used in this thesis are shown in the *Appendix 1*.

2.7.4 Bisulphite primer design

Promoter sequences, obtained from the Ensembl genome browser, were uploaded and in *silico* bisulphite converted. Primers for bisulphite analysis were then designed with the Bisearch Primer Design and Search Tool (http://bisearch.enzim.hu/) using default settings. Sequential nested primer sets were selected based on the threshold score, which takes into account multiple amplification efficiency variables, and on the product size and genomic location of primer sets. Bisulphite primer sets are shown in *Appendix 1.*

Chapter 3

Identification of candidate genes regulated by DNA methylation

3.1 Introduction

The precise functional roles of DNA methylation in mammals are far from determined. The current dogma predicts methylation of CpG dinucleotides silences some repetitive elements (Walsh et al., 1998), has an important but secondary role in X-inactivation (Lock et al., 1987; Heard et al., 1997) and appears to be the primary mark directing mono-allelic expression of embryonic imprinted genes (Murrell et al., 2004; Bartolomei, 2009). However, the extent to which promoter methylation is the primary mechanism of repression at single-copy loci is unknown. While many studies have demonstrated correlative relationships between gene expression and CpG methylation, currently there are no examples in the literature of genes directly and unequivocally regulated by DNA methylation *in vivo*.

As part of a strategy to identify and characterise genes regulated by DNA methylation, an initial set of candidate loci must be derived. The fundamental aim of this chapter is thus to identify novel candidate genes putatively regulated by promoter CpG methylation. These genes will then be used as targets for comprehensive cause and effect analyses in subsequent chapters. The method used to derive candidate genes here is of crucial importance. A balance must be struck between inclusion and specificity to eliminate false-positives yet include what could potentially be a very small number of *bona fide* methylation-dependent genes. Previous studies have identified candidates by screening for genes that are differentially methylated during development (Maatouk et al., 2006; Ng et al., 2008) or tissue-specifically (Suzuki et al., 2007; Shen et al., 2007; Weber et al., 2007;). However, it is not clear in these reports whether the observed changes in DNA

methylation are directed by other upstream epigenetic mechanisms or indeed, whether differential methylation is a cause or a consequence of transcriptional activity. Moreover, due to the restricted number of tissues or time points that can be tested, only a limited number of candidates can be identified. Another approach to identify methylation-dependent candidate genes is to examine the expression profile of ES (Fouse et al., 2008) or somatic cells (Jackson-Grusby et al., 2001) lacking DNA methyltransferases. This method identifies a large number of mis-expressed target genes but the majority would be predicted to be false-positives or indirect hits as a result of global changes in transcription factor networks and/or other epigenetic modifications. This hypothesis is borne out through *in silico* analysis of the upregulated genes, which demonstrates ~75% are associated with CpG island promoters which are unmethylated prior to Dnmt1 inactivation. It is therefore unclear which, if any, of the genes upregulated in genetic studies would be candidates for direct regulation by promoter DNA methylation *per se*.

In this chapter I have used a novel combination of genetic and biochemical analyses of gene expression patterns to identify candidate genes regulated by DNA methylation. By cross-referencing candidate genes from three independent experimental approaches, this method enriches for targets that consistently respond to different functional assays that test the role of DNA methylation in maintaining gene silencing. This combination of complementary approaches is predicted to reduce indirect hits and false-positives, and therefore select for *bona fide* candidate genes. In summary, through requiring candidate genes to satisfy several successive experimental predictions and employing highly stringent thresholds, I have generated a novel set of candidate genes potentially regulated by promoter CpG methylation for downstream analysis in future chapters.

3.2 Mouse Embryonic Fibroblasts lacking Dnmt1 are globally hypomethylated

As the first experimental approach to identify candidate loci I utilised *Dnmt1, Trp53* double-null mouse embryonic fibroblasts (MEFs) that have been reported to be

globally hypomethylated (Lande-Diner et al., 2006). Because the absence of Dnmt1 in somatic cells induces p53-mediated apoptosis it is necessary to inactivate p53 (encoded by *Trp53*) in these MEFs. To control for this, MEFs lacking only p53 were used (both cell lines were a kind gift of Prof. Howard Cedar). For brevity, the hypomethylated *Dnmt1*$^{-/-}$ *Trp53*$^{-/-}$ cells are hereafter referred to as the DP line and control *Trp53*$^{-/-}$ (*Dnmt1*$^{+/-}$) cells are referred to as the P line. I predicted that genes de-repressed in DP cells would form the initial broad set of candidate loci for subsequent analyses to refine to an enriched list. Prior to a global gene expression analysis, I characterised and confirmed the genotype and cellular properties of DP and control P cells

The knockout strategy used to generate the DP line deleted a 900bp region at the 5' end of the somatic Dnmt1 transcript (Li et al., 1992). This corresponds to deletion of 20bp of the 5'UTR, the coding portion of the first exon and the splice acceptor site (*Fig 3.1*). Because this deletion truncates the N-terminal region of Dnmt1 protein it is known as the *n*-allele, to distinguish it from the subsequently generated *c*-allele, which disrupts a portion of the C-terminal domain (Lei et al., 1996). To confirm DP cells lacked Dnmt1 I performed quantitative RT-PCR (qRT-PCR) to determine whether Dnmt1 transcripts were present. Using primers that overlapped the 5' deleted region, I was unable to detect Dnmt1 transcript in DP cells but could readily detect Dnmt1 in control P cells (*Fig 3.1 left panel*). This data suggests that as expected DP cells lack the 5' region of Dnmt1 *(n*-allele*).* However, qRT-PCR directed to the 3' region of the transcript showed Dnmt1 RNA was present in DP cells but at 7-fold reduced levels compared to P cells (*Fig 3.1* right panel). This data is consistent with previous publications, which have noted that alternative splicing of the *n*-allele leads to a truncated but functional Dnmt1 protein albeit at significantly reduced levels (Lei et al., 1996). Importantly, at these reduced levels the abundance of Dnmt1 is not sufficient to maintain global DNA methylation levels and cells still become progressively hypomethylated over just a few generations.

To confirm the DP cells used here were hypomethylated, I used the 5mC antibody, which recognises methylated cytosines. Staining with the 5mC antibody was

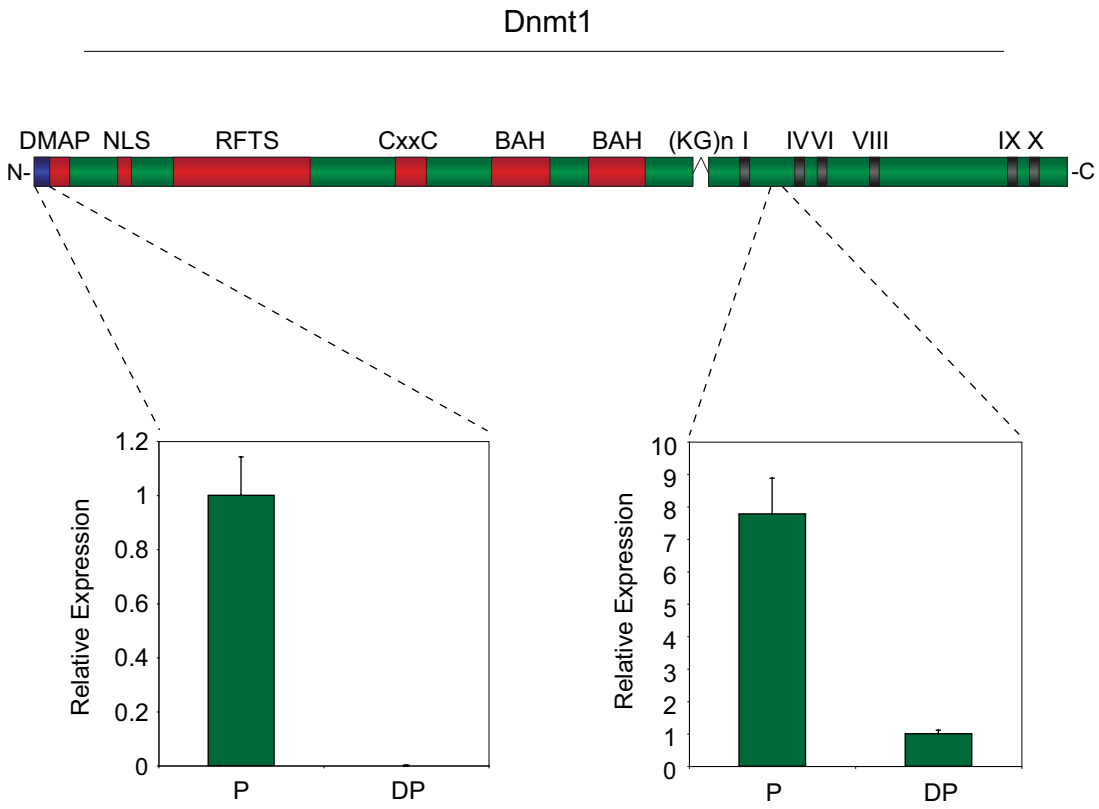


Fig 3.1. Analysis of Dnmt1 transcript levels in DP and P cells. Top: schematic of mouse Dnmt1 protein (1616 amino acids) showing the N-terminal region deleted in DP cells (shown in blue) (n-allele) and the functional domains (red) . The N-terminal domain of Dnmt1 contains a nuclear localization signal (NLS), an CXXC DNA-binding motif, a bromo-adjacent homology domain (BAH), a proliferating cell nuclear antigen-binding domain (PBD) that contributes to targeting to the replication fork (*see Fig 1.4*) and replication foci targeting sequence (RFTS). The C-terminal catalytic domains includes six conserved motifs shown in black: Lower left panel: qRT-PCR demonstrating Dnmt1 transcript was not detected in DP cells using primers that overlapped the deleted region, indicating these are bona fide n-allele Dnmt1-null cells. Lower right panel: qRT-PCR showing primers directed toward the C-terminal region could detect weak Dnmt1 transcript in DP cells suggesting an alternatively spliced transcript was being produced. Shown is relative expression normalised to Gapdh.

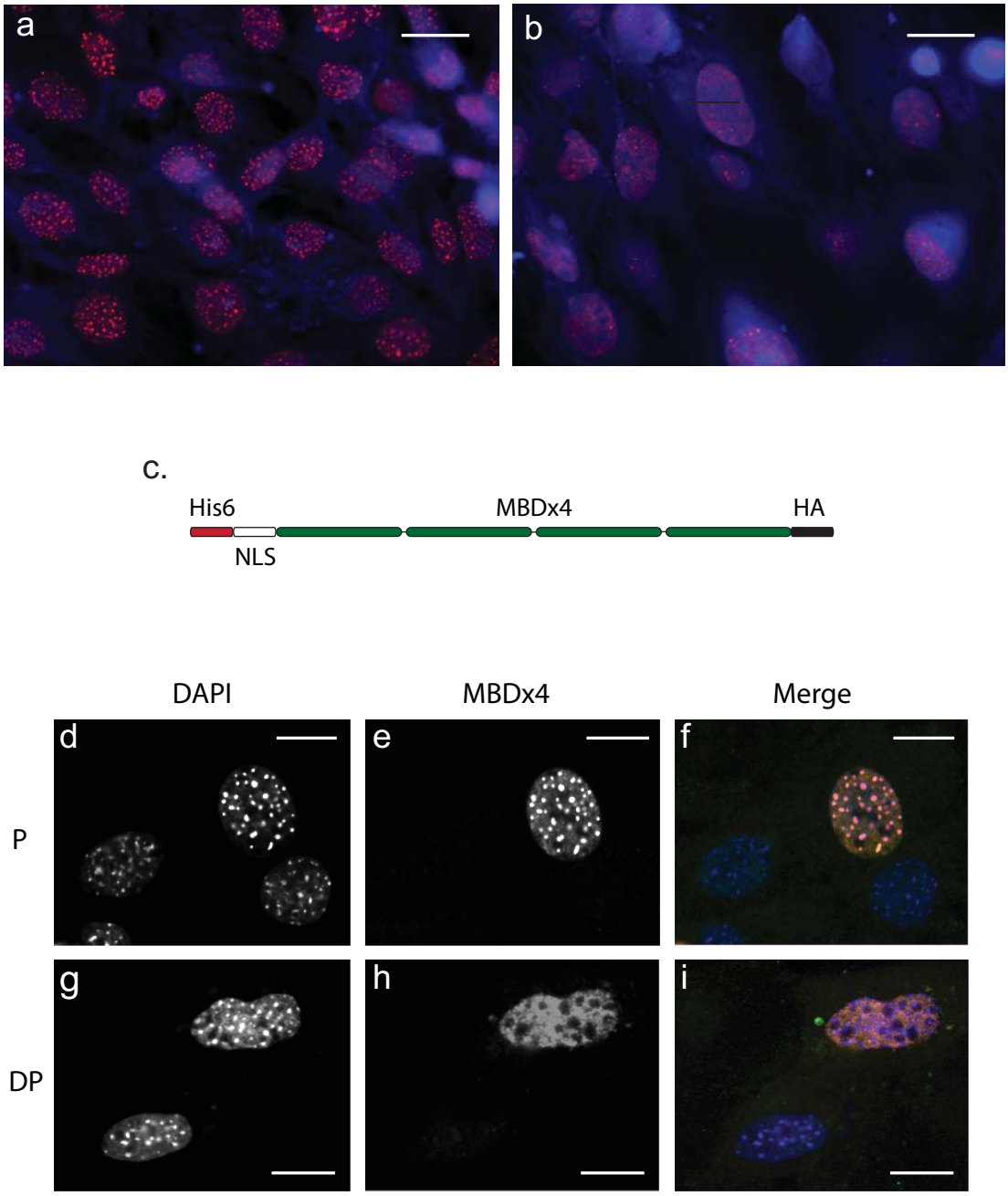


Fig 3.2. **DP cells are globally hypomethylated.** a.) P cells and b.) DP cells were fixed and stained with the 5mC antibody (red) and DAPI (blue). a.) P cells exhibited strong staining at discrete foci whereas b.) DP cells showed significantly reduced and diffusely distributed staining indicating DP cells were globally hypomethylated. Scale = 20μm. c.) Structure of the MBDx4 protein which binds methylated CpGs with high avidity (Jorgensen et al., 2006). d & g.) P and DP cells were stained with DAPI and e & h.) transfected with the MBDx4 expression construct and stained with HA antibody. f & i.) Merge showing strong DNA methylation (red) co-localises with heterochromatin chromocentres (blue) (DAPI) in P cells but is weakly present, diffusely distributed and exclusive of chromocentres in DP cells. Scale = 10μm.

markedly reduced in DP cells compared to P cells (*Fig 3.2a & b*). Furthermore, the residual methyl-cytosine staining detected in DP cells was diffusely dispersed around the nucleus (*Fig 3.2b*). In contrast the strong methyl-cytosine staining in P cells formed distinct nuclear foci, which probably represent heterochromatic chromocentres (*Fig 3.2a*). To ascertain whether residual DNA methylation in DP cells was mis-localised because a loss of heterochromatin or due to alternative reasons, such as reiterative *de novo* methylation, I used the poly-MBD method developed by Jorgensen et al (2006). Here, a tagged poly-MBD1 methyl-binding domain fusion protein (poly-MBDx4) is transiently expressed in cells (*Fig 3.2c*). This protein binds methylated CpGs with ~100-fold higher avidity than the 5mC antibody. Crucially, unlike with the 5mC antibody, there is no requirement to denature cellular DNA with this method. This allows a comparison of the distribution of residual DNA methylation with native heterochromatin. Transfection and analysis of the poly-MBDx4 protein showed that residual DNA methylation in DP cells is localised to non-heterochromatic regions of the genome (*Fig 3.2f*). In contrast, staining in P cells strongly overlapped with DAPI bright spots (*Fig 3.2i*). This suggests that in hypomethylated somatic cells (DP) chromatin remains capable of compacting into heterochromatic structures, as has been noted in ES cells, but that the low levels of residual methylation no longer localises to heterochromatin (Gilbert et al., 2007). It is noteworthy that the apparently high background level of methylation in DP cells is a consequence of the high avidity of the fusion protein used here combined with diffusely distributed residual methylation and not due to intrinsically elevated methylation levels in DP cells. Indeed, mass spectrometry (MS) comparison of genome-wide methyl-cytosine levels indicated early passage DP cells retained methylation at ~40% of the level of P cells. This is consistent with a previous study that reports Dnmt1-null (*n*-allele) somatic cells (DP) maintain ~35% global methylation levels (Lei et al., 1996).

To confirm P cells were comparably methylated to primary cell types I digested DNA from brain tissues, P cells and DP cells with methyl-sensitive enzymes. Here, ethidium bromide staining of *MaeII* (methyl-sensitive) digested genomic DNA suggested that P cells had similar global methylation levels to primary MEFs

(pMEF) whereas DP cells appeared significantly hypomethylated (*Appendix 2*). I further evaluated DNA methylation at minor satellite repeats by Southern blot. Here, genomic DNA extracted from the DP line was susceptible to digestion whereas DNA from the P line was sensitive to *MspI* but crucially, was resistant to digestion by methyl-sensitive *HpaII* and *CfoI* enzymes suggesting minor satellite sequences are fully methylated in the P line but hypomethylated in DP cells (*Appendix 2*). I conclude that DP cells are globally hypomethylated whereas P cells retain normal genome-wide DNA methylation levels.

3.3 DP and P cells proliferate and are non-transformed

All DP and P cells used in this thesis were male cells between passages p5 to p11. Following a short lag after their derivation (p1-p3), both DP and P cells grow at normal exponential rates and do not undergo senescence, even after 100 generations (Lande-Diner et al., 2006). To investigate both the proliferative capacity and the state of transformation of DP and P cells used here, I initially cultured both cell lines on *Poly*-2-hydroxyethyl methacrylate (poly(HEMA)) coated surfaces, which inhibit cell adhesion (Fukazawa et al., 1996). Because anchorage-independent growth is a property of transformed cells, proliferation in the presence of poly(HEMA) indicates cells are transformed (Shin et al., 1975). To measure cellular growth rates I used a colourimetric assay whereby yellow MTT is reduced to purple formazon in living cells. This can be accurately measured in microplates by a spectrophotometer. This analysis demonstrated that neither DP nor P cells (p9) were able to significantly proliferate on poly(HEMA) coated plates in comparison with control transformed 293T cells or with non-coated surfaces (*Fig 3.3a & b*). This suggests that despite the absence of p53 these MEFs (DP & P) have not undergone the transition to a transformed phenotype and represent a relatively primary cell-type. Additionally, analysis of the control non-coated growth rates suggested that DP cells proliferated at ~75% the rate of P cells, until contact inhibition slowed cellular division (day5) (*Fig 3.3a*). This is consistent with second-order proportionality of rRNA synthesis in hypomethylated cells (Maaloe & Kjeldgaard, 1966; Gagnon-Kugler et al., 2009). To confirm the growth properties of DP and P cells I conducted a soft agar assay. Here

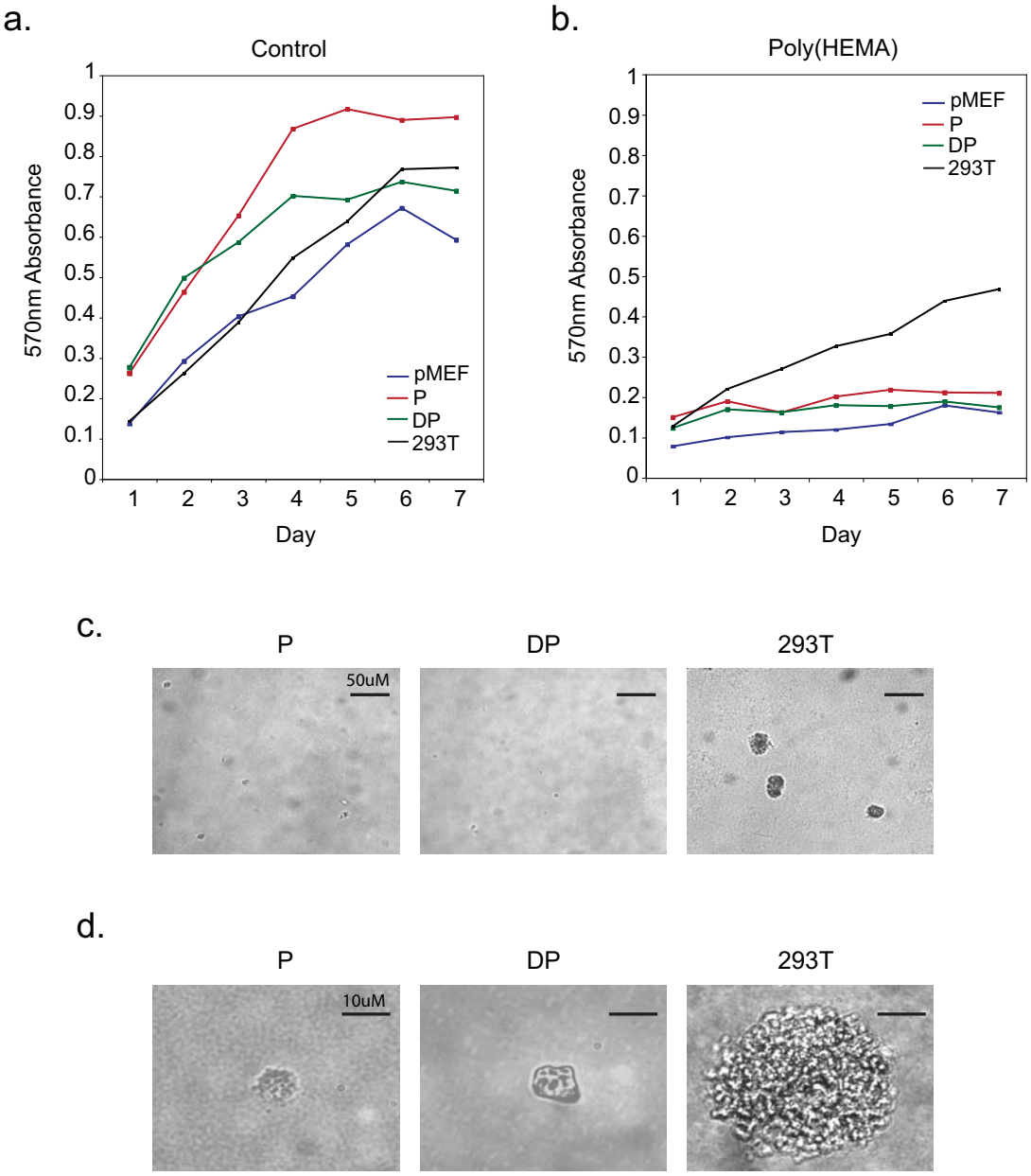


***Fig. 3.3*. DP and P cells are non-transformed.** Cells were plated at low density on a.) control non-coated or b.) poly(HEMA) coated microplates. a.) All cell lines proliferated at similar rates until maximum density at ~day 4-5. b.) Only transformed 293T cells were capable of proliferation on poly(HEMA) indicating DP and P cells are non-transformed. c & d.) Cells were seeded at low density in an agar/growth medium matrix and cultured for 12 days to determine their state of cellular transformation. Anchorage independent colonies were only observed for positive control 293T cells shown at c.) low and d.) high magnification. Scale bars c.) 50uM and d.) 10uM.

DP, P and positive control 293T cells were seeded in low-density agar/growth media matrices and cultured for 12 days (*Fig 3.3c & d*). Only 293T cells were able to form anchorage-independent colonies consistent with the conclusion that both the DP and P cells used in this thesis are not transformed.

3.4 Candidate gene identification principal

The fundamental aim of this chapter is to identify candidate genes regulated primarily by DNA methylation. To do this, I resolved that candidate targets must pass several experimental criteria. Thus, potential candidate genes will be filtered through three successive experiments, which progressively deplete false positives and indirect hits. Genes that are positive in all three experiments represent my candidate targets and can be analysed further. As such, the experimental criteria for candidate genes are thus:

1. They must be de-repressed in hypomethylated DP cells (*Section 3.5*).
2. They must be de-repressed in pMEFs and P cells treated with 5-aza deoxycytosine (5-aza dC) (*Section 3.6*).
3. Following treatment with 5-aza dC, candidate genes must continue to be aberrantly expressed after 14 days recovery in the absence of 5-aza dC (*Section 3.7*).

This last criterion is crucial. Genes that turn 'on', in response to the demethylation inducing nucleoside analogue 5-aza dC, may be affected indirectly (for example as part of an apoptotic response) or may be additionally marked by alternative epigenetic modifications. Both these classes of target gene would be predicted to re-impose silencing during a 14 day recovery period either through a return to normal cellular physiology or through epigenetic marks re-targeting gene silencing (including re-targeting DNA methylation), respectively. However, genes that rely exclusively on DNA methylation for epigenetic memory would be predicted to remain expressed even after 14 days recovery, as there would be no mechanism of reiterative DNA methylation targeting in somatic cells. Thus, this novel approach

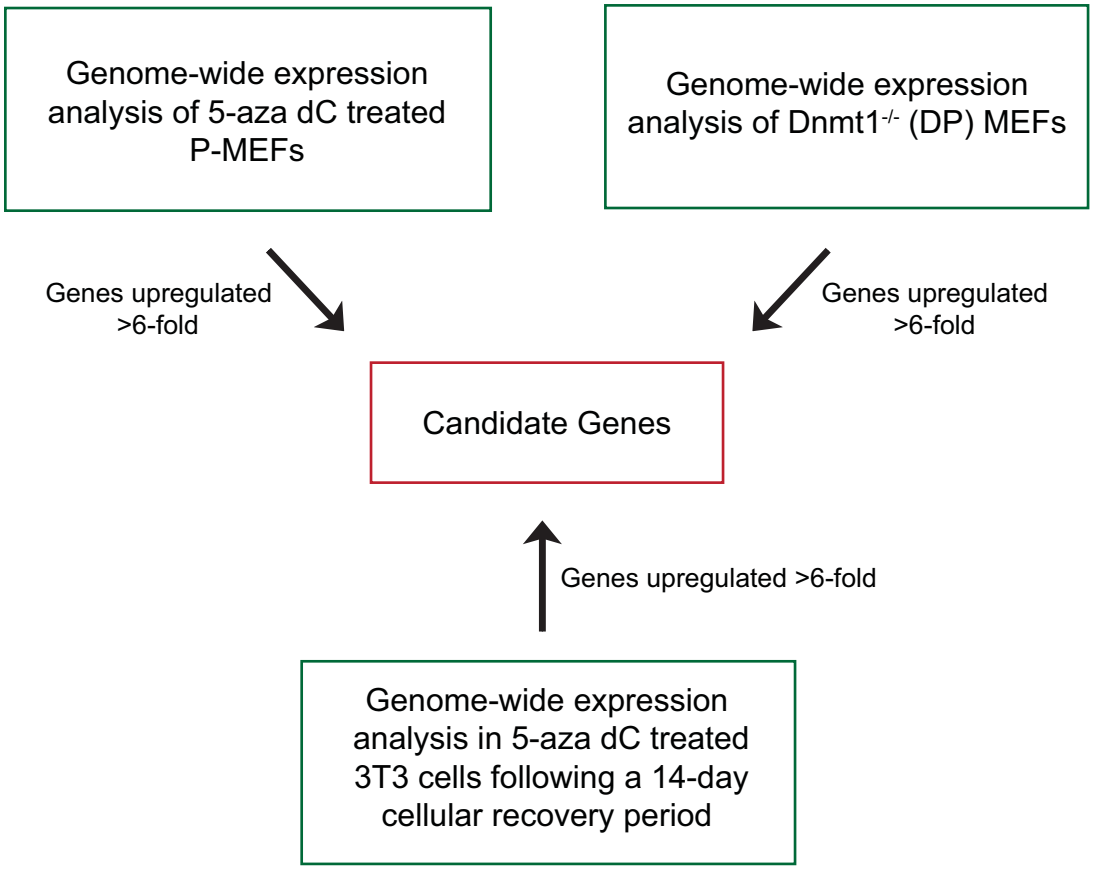


Fig. 3.4. **Schematic for identification of candidate genes regulated by DNA methylation.** Genes must be upregulated >6-fold in all three genome-wide experiments to be considered a methylation-dependent candidate gene.

cross-referenced with experiments 1 and 2 should identify *bona fide* candidate genes potentially regulated by DNA methylation for further analysis (*Fig 3.4*).

To evaluate global gene expression levels in each experiment I used Illumina Mouse Ref-8 v2 or Mouse Ref-6 v2 Expression BeadChips. The Ref-8 chips contain probes for 25,697 transcripts (18,098 genes) from the mouse NCBI RefSeq database whereas the Ref-6 chips include these probes and an additional 19,584 probes that correspond to various gene isoforms and repeat elements. A crucial aspect of this process is where to set threshold limits that distinguish significant changes in gene expression between samples. I initially used a t-test with average normalisation, applying multiple testing corrections using Benjamini and Hochberg false discovery rate. However, because of the relatively high number of biological and technical replicates used, this method would often report expression changes as low as 1.2-fold were highly statistically significant (p = <0.01). Biologically, genuine targets regulated by DNA methylation would be predicted to change expression considerably in the absence of promoter methylation. This is because promoter methylation is proposed to promote strong gene silencing. If DNA methylation is the primary system for regulating expression, depletion should result in robust gene activation (i.e. from nothing to something) and therefore a significant net change in expression. To account for the discrepancy between what is statistically significant and biologically significant with respect to expression changes according to DNA methylation, I resolved to use a highly stringent 6-fold expression change threshold. This strict threshold is expected to minimise false-positives and maximise the pool of genes crucially reliant on DNA methylation patterns for their expression. In contrast, previous studies investigating regulation by DNA methylation have set their threshold at 2-fold (Jackson-Grusby et al., 2001; Fouse et al., 2008). While this undoubtedly increases the number of potential genuine targets, it makes identification of the direct targets within a large dataset very difficult.

3.5 Germline specific genes are de-repressed in DP cells

To identify candidate genes de-repressed in the absence of DNA methylation I initially assayed global gene expression levels in DP and control P cells using

Illumina Mouse Ref-8 BeadChips. To validate the array method I initially plotted biological replicates of DP or P hybridisations. Here, DP cRNA replicates showed a strong correlation (pairwise r^2 = 0.99), as did P cell cRNA biological replicates (pairwise r^2 = 0.98) (*Fig 3.5a & b*) and the data was distributed as expected with 47.5% (DP) and 48.0% (P) of genes considered to be expressed (p = <0.05), respectively (*Fig 3.5c*). The distribution of data and the strong replicate correlations indicated that RNA labelling and hybridisation produced accurate and reproducible data. Comparison of the DP and P datasets showed a correlation of r^2 = 0.94 (*Fig 3.6a*). Using a stringent 6-fold relative expression threshold I identified 221 genes upregulated in DP cells and 171 genes downregulated. This corresponds to 1.2% and 0.9% of the array, respectively (*Fig 3.6b*). As expected, more genes were upregulated than downregulated, consistent with the proposed role of DNA methylation in gene silencing (Bird, 2002). However, a significant number of genes *were* downregulated in hypomethylated DP cells, potentially due to changes in transcription factor availability or epigenetic cascades. Interestingly, many expression changes (down & up) must be stochastic or indirect, as comparison of the array data generated here with a previously published expression analysis of hypomethylated fibroblasts (Lande-Diner et al., 2006) showed only a modest correlation. Indeed, I was unable to validate several of the top hits reported in this study (data not shown) (Lande-Diner et al., 2006). This may reflect the fact that most expression changes in hypomethylated cells occur independently of direct changes in DNA methylation and therefore are under multiple, partially stochastic, influences. Identifying the potential methylation-dependent targets within this broad DP cell dataset is the key aim of this chapter (*Fig 3.4*).

I validated the microarray expression data through RT-PCR of targets the array identified as upregulated >6-fold in DP cells (*Dazl, Slpi, IAP* elements, *Tex13*) or as showing no significant expression change (*Gata6, Krt8, Crip1)* (*Fig 3.7a*). Additionally, as a negative control I confirmed that ES-cell specific transcripts were not expressed in either cell line, as these genes have been reported to be regulated by multiple tiers of epigenetic repression and are not re-activated by demethylation (*Fig 3.7b*) (Feldman et al., 2006; Cedar & Bergman, 2009). My RT-PCR analysis was

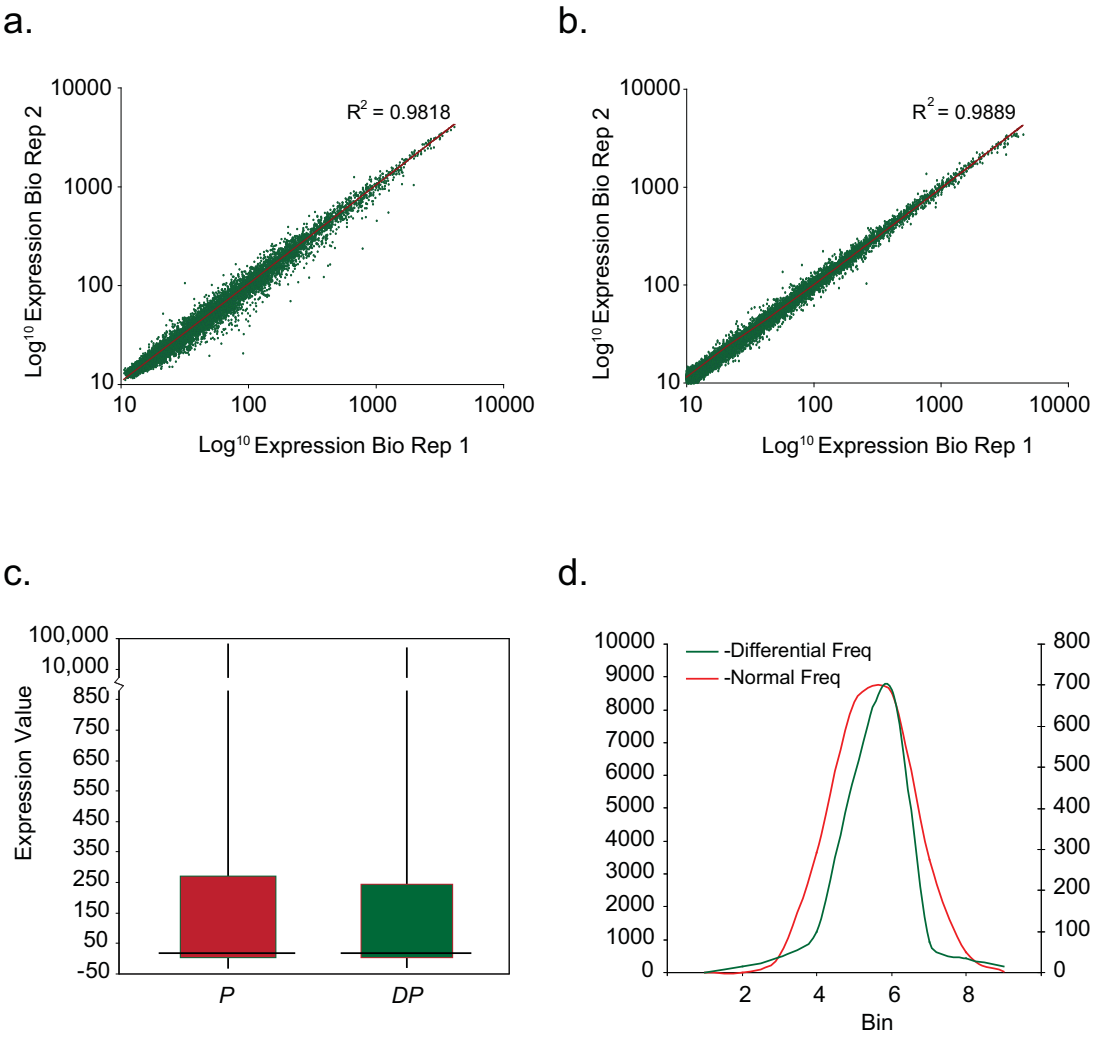
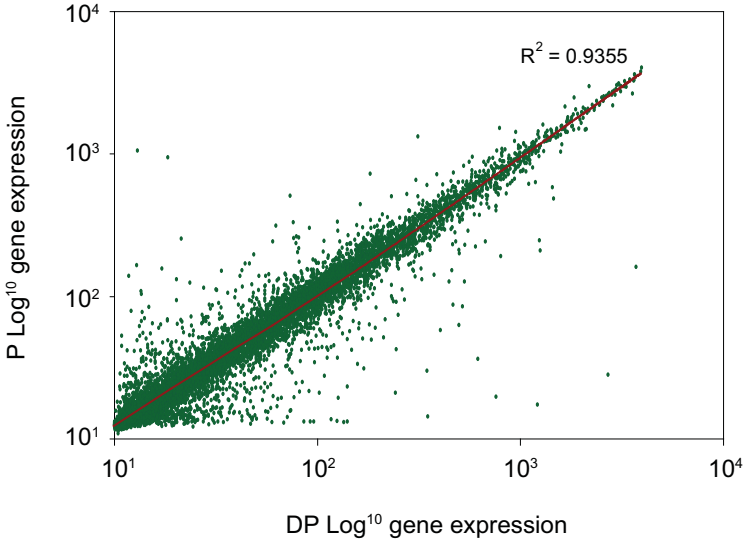


Fig 3.5. **Microarray quality control.** Log^{10} expression data points for a.) P and b.) DP microarray biological replicates were plotted to asses the reproducibility of the array and biological variation inherent within different time points. Both datasets showed strong correlation indicating there was little variation between biological replicates of the same sample. c.) Box blot showing the distribution of raw data values between P and DP samples is comparable. Each box represents the 25^{th} and 75^{th} percentile with the horizontal line the median value. Whiskers show the 1^{st} and 99^{th} percetile d.) Raw data from DP cells (green) is normally distibuted with a slight shift to the right indicating more genes are upregulated than downregulated. Red line shows expected normal distribution.

a.



b.

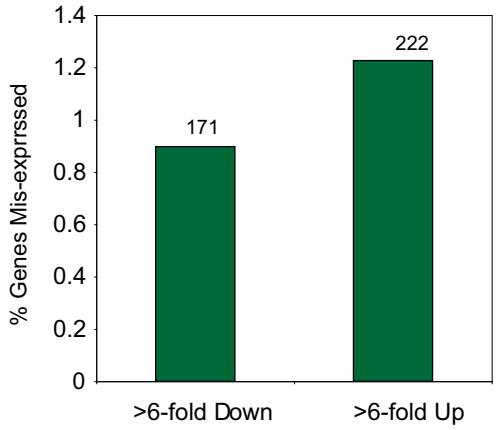


Fig 3.6. **Microarray expression analysis of DP and P cells.** a.) Scatter plot analysis of P vs DP Log^{10} expression values. b.) Number of genes mis-regulated >6-fold in microarray analysis of DP cells relative to P cells.

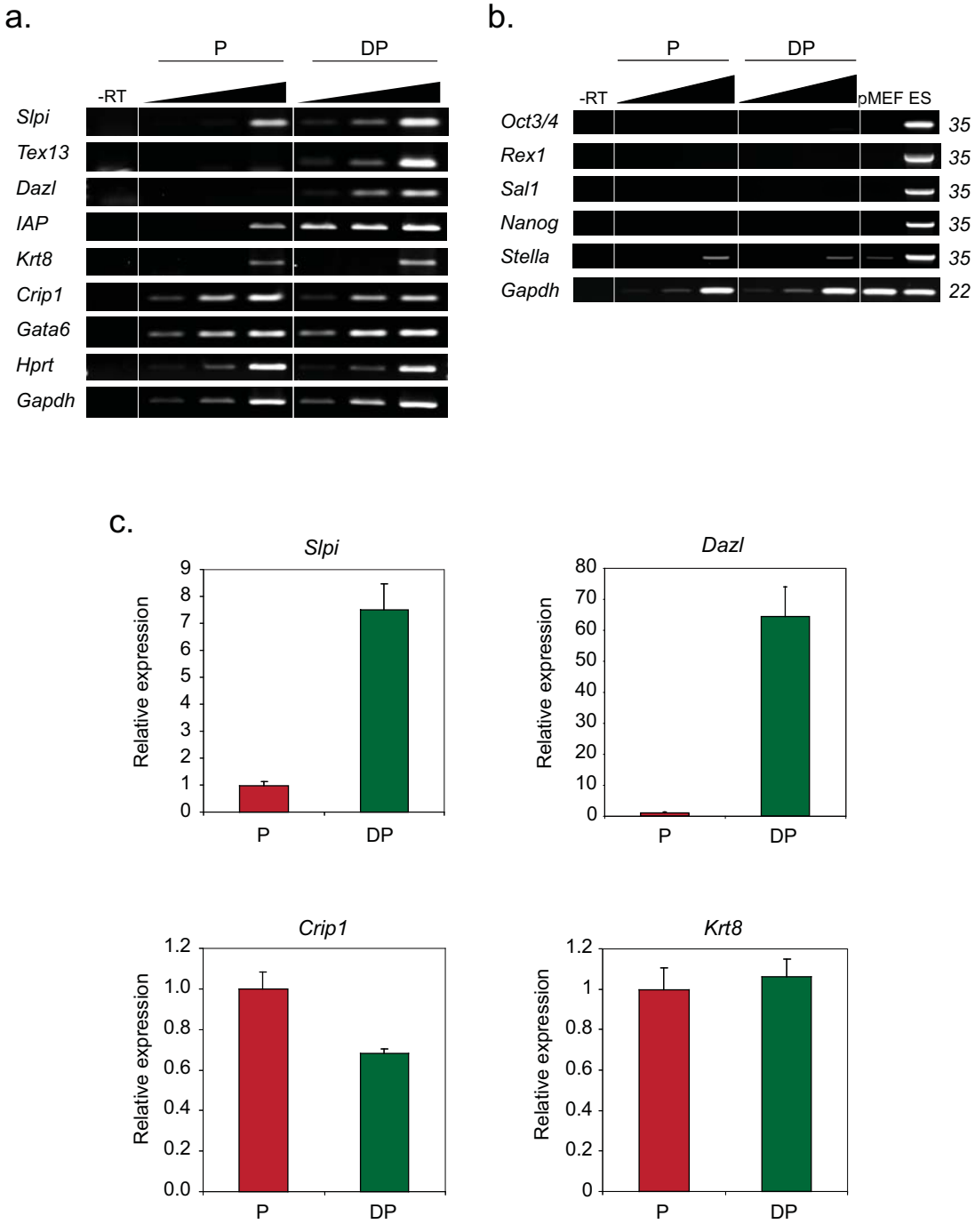


Fig 3.7. **Validation of expression microarray results.** a.) Semi quantitative RT-PCR validation of genes predicted to be upregulated in DP cells (*Slpi, Tex13, Dazl, IAP*) or to exhibit no change (*Krt8, Crip1, Gata6*). cDNA dilutions are 1/50, 1/20 and 1x. *Hprt* and *Gapdh* are loading controls b.) Negative control semi-quantative RT-PCR of ES-cell specific transcripts. c.) qRT-PCR confirmation of relative gene expression. Shown is expression relative to P cells and normalised to *Gapdh.* Error bars represent propergated S.E.M.

highly correlative with the array predictions, with all 12 transcripts tested following the expected expression pattern. Furthermore, relative fold-changes in expression reported by the array were comparable to quantitative RT-PCR (qRT-PCR) validation (*Fig 3.7c*). However, for large expression changes, the array analysis often underestimated the relative fold change. This probably reflects the limited dynamic range of Illumina BeadChips but does not significantly effect this analysis, as relative discrepancies were only observed with changes >10-fold, beyond the 6-fold threshold used here. These validations indicate that this array platform provides high quality reproducible data for further analysis. Because this study is focused on genes regulated by DNA methylation in *cis*, only upregulated genes are considered from here. A comprehensive list of all targets upregulated >6-fold is shown in *Table 3.1*.

Gene	Chromosome	Function	Fold change
Tex19	**11**	**Retrotransposon silencing**	**430.1**
Fmr1nb	X		317.1
Rhox5		**Homeobox transcription factor**	**306.2**
Casp1	9	Caspase cascade in apoptosis	304.5
Nckap1l		Regulation of actin cytoskeleton,	280.2
Slc47a1	**11**	**Ion transport**	**260.0**
Fkbp6	5		208.1
Magea2	X		198.6
OTTMUSG0000010673	4		182.8
Psg23	7	Immunoglobulin	181.0
Gdf15	8	Growth factor	139.7
Plac8	5		120.5
Pet2	X		118.8
Slpi	2	Leukocyte peptidase inhibitor	111.5
Krt14	11	Intermediate filament protein	95.6
MageK1	**X**	**Mage protein**	**93.1**
Hamp2	7	Hepcidin antimicrobial peptide	89.3
EG434729	X	Iron storage	87.8
Taf7l	X	Basal transcription factor	87.2
Pkd1l2	8		76.3
Pira4	7		68.9
Tex13	**X**		**60.2**
Gpr97	**8**	**G-protein coupled receptor**	**56.3**
Gm773	X		51.4
Rpl39l	16	Ribosomal protein	48.0
Nlrp4c	7		47.3
Mrvi1	7		46.4
2010005H15Rik	16		43.4
Amhr2	15	Cytokine-cytokine receptor interaction	40.5
Msln	17	Cell adhesion	39.8
Mov10l1		**Folate biosynthesis**	**39.1**
Apol9b	15	Lipid binding,	36.7

Pde6h	6		36.7
Xdh	17		35.5
Dpt	1		34.8
Magea5	X		33.7
1700034E13Rik	18		33.1
OTTMUSG00000015743	2		31.9
Jam4	16	Cell adhesion	31.5
Tnfrsf26			30.6
Rps4y2			**30.0**
Usp18	6	Protease	29.2
LOC100048346			28.5
Acpp	9	Riboflavin metabolism	28.4
A030004J04Rik	3		28.2
Foxq1	13	Fork head transcription factor	27.9
Asz1	6	Sam transcription factor	27.8
Prl	13	Somatotropin hormone	27.7
Gstp2	**19**	**Glutathione metabolism**	**27.2**
Tex11	X		26.8
Pvalb	15		26.7
V1rb3	6		26.4
D12Ertd647e	12		26.0
EG433016	16		25.7
Aim2	1		25.4
Parp14	16		25.1
Selp	1	Immune system process	24.9
Iap		**Reteroelement**	**24.4**
Rem1	2		23.6
Serpinb6b	13	Protease inhibitor	22.9
V1rd21	7	Vomeronasal receptor	21.5
AA467197	2		21.5
LOC547343	17	Class i histocompatibility antigen,	21.2
Rhox4d	**X**	**Homeobox transcription factor**	**21.1**
Tph2	10	Tryptophan metabolism	20.5
Ugt1a10	1	Androgen and estrogen metabolism	20.5
Xlr4a	X		19.7
Piwil2	**14**	**Argonaute and dicer protein**	**19.7**
Pira6			19.5
Ctsk	3	Peptidase	19.5
Oasl2	5		19.3
Olfr1463	19	Olfactory receptor	19.2
Sh3tc2	18		19.0
Cyp2b13	7	Cytochrome P450	18.9
Akr1c12	13		18.7
B3gnt3	8	Glycosphingolipid biosynthesis	18.7
Trim30	7		18.6
Dpep1	8	Membrane dipeptidase	17.9
Hspb2		:Alpha-crystallin-related small heat shock protein	17.8
Nr1h4	10	Regulation of cholesterol metabolism	17.6
Gsta1	9	Glutathione metabolism	17.6
Pla1a			17.4
AI467606	7		17.2
LOC100048710			17.1
Ccdc114	7		16.7
Xist	X	X-Chromosome inactivation	16.3

LOC100038882	4		16.2
Efhc2	X		16.0
Cidec	6	Caspase-activated nuclease	15.4
Ang2	14		15.3
Nppb	4	Natriuretic peptide	15.1
Lpin3	2		15.1
Kcnab2	4		15.0
Il15	8	Cytokine-cytokine receptor interaction	14.8
Krt7	15	Intermediate filament protein	14.6
Tns4	11		14.1
Tap2	17	Antigen processing and presentation,	14.0
Cldn15	5		13.6
Tex19.2	**11**		**13.5**
Mb	15	Myoglobin, metal ion-binding site	13.5
Fyb	15	Fyn binding protein	13.5
Defb8	8	Beta defensin,	13.3
Gsta2	9	Glutathione metabolism	13.1
Adamts2	11		13.0
Pstpip1	9		12.7
Itgb7	15	Cell adhesion	12.6
Dnase1l3		Deoxyribonuclease	12.4
Hoxc13	15	Homeobox transcription factor	12.4
Hspb7		Alpha-crystallin-related small heat shock protein,	12.4
D6Mm5e	6		12.2
Arl11	14	Adp-ribosylation	12.2
Hoxc10	15	Homeobox transcription factor	12.1
Rasl12	9	Ras gtpase	12.1
Tnnt1	7	Troponin	12.0
Slc15a3	19	Amino acid transport and metabolism,	11.9
Plch2	4	Lipid metabolic process	11.7
Lamb3	1	Cell communication, Extracellular matrix structural constituent	11.7
Serpinb9g	13	Protease inhibitor	11.6
Casp4	9	Induction of apoptosis	11.6
Was	X	Regulation of actin cytoskeleton	11.6
LOC236749	X		11.5
Hoxd10	2	Homeobox transcription factor	11.3
AU022751	X		11.3
Tnfsf13b	8	Cytokine-cytokine receptor interaction	11.1
Slc12a8	16		11.1
Dnmt3l	10	Methyltransferase cofactor	11.0
4930481M05	X	Somatotropin hormone	10.9
Eng	2		10.7
Mylc2pl			10.6
Pcdhb3	18	Cell adhesion	10.2
Uroc1	6	Histidine metabolism	10.1
Aldh3a1	11	Aldehyde metabolic process	10.0
Slc12a1	2	Na-k-cl co-transporter,	9.7
Crip2	12		9.6
Rsad2	12		9.6
Cdsn	17		9.5
LOC100038908	7		9.4
Prl2c4	13		9.4
Plscr2	9	Scramblase	9.3

Bst2	8	Bone marrow stromal antigen, Cell communication	9.2
Acsbg1	9	Amp-dependent synthetase and ligase	9.2
LOC631002	X		9.2
Tex14	11		9.1
Arhgap30	1		9.1
Acta1	8	Regulation of actin cytoskeleton	8.9
BB146404	10		8.8
Prl2c3	13	Somatotropin hormone	8.8
Prelp	1	Class ii small leucine-rich proteoglycan	8.7
P2ry14	3	Neuroactive ligand-receptor interaction	8.7
Mnd1	3		8.7
Tnnt3	7	Troponin	8.6
Gpr114	8	G-protein coupled receptor	8.5
EG667977	17		8.4
Rapsn	2		8.4
Chrng	1		8.2
Cntn3			8.2
Fkhl18	2	Fork head transcription factor	8.1
Abi3			8.0
Ttn			8.0
EG630499	17		8.0
Havcr2	11		8.0
Timm8a2	14		8.0
Iigp2	11	Interferon-inducible gtpase	8.0
Gatm		Urea cycle and metabolism of amino groups	8.0
Ly6c1	15		7.9
AI747699			7.9
Fbln7		Calcium ion binding	7.9
9030224M15Rik	10		7.6
Slc14a1	18		7.6
Hao1	2	Glyoxylate and dicarboxylate metabolism	7.6
Bcar3	3		7.6
Cgnl1	9	Myosin tail	7.5
Mgst2	3		7.5
Uts2r	11	Neuroactive ligand-receptor interaction,	7.4
Tnnc2	2	Calcium signaling pathway	7.4
Tcfl5	2	Hlh transcription factor	7.4
Afp	5	Serum albumin	7.3
Cyp4a12a	4	Cytochrome P450	7.3
LOC100044314			7.2
Mylpf	7	Regulation of actin cytoskeleton	7.2
Psp	2	Lipid-binding serum glycoprotein	7.2
Lgals3bp	11		7.1
LOC100048554			7.1
Evi2a	11		7.0
1700012B09Rik	9		7.0
Myod1	7		6.9
Aqp1		Aquaporin,	6.9
Cct6b	11	Molecular chaperone	6.9
Angpt4	2		6.8
9130218O11Rik	15		6.8
Psmb9	17	Multicatalytic endopeptidase	6.8
Cox8b	7	Oxidative phosphorylation,	6.7
Arhgap9	10	Gtpase	6.7

Tuba3a	6	Cytoskeleton component	6.7
Isg20	7		6.7
EG638695	14		6.7
Samd9l	6		6.7
Soat2	15	Lipid metabolism	6.6
Cck	9	Gastrin/cholecystokinin peptide hormone	6.6
Mfap5	6		6.6
1500015O10Rik	1		6.6
S3-12	17		6.5
1700112C13Rik	9	Protease	6.5
Tor3a			6.5
EG633640	13		6.5
Slc22a4	11	Ion transport	6.5
Piwil4	9	Argonaute and dicer protein	6.4
Angptl4	17	Ppar signaling pathway, angiogenesis	6.4
D10Bwg1379e	10		6.4
Hc	2	Complement and coagulation cascades	6.3
Pip5k1b	19	Inositol phosphate metabolism	6.3
Opn1mw	X	Photoreceptor	6.3
Anxa8	14	Blood coagulation	6.3
AI451557	8		6.3
BC010462		Immunoglobulin subtype	6.2
Actg2	6	Muscle protein	6.2
Mlkl	8		6.2
Mmp17	5		6.1
Ccl17	8	Cytokine-cytokine receptor interaction,	6.1
H19	7		6.0
Rpl3l		Ribosomal protein	6.0

Table 3.1. **Transcripts significantly de-repressed in DP cells.** Shown are all transcripts de-repressed >6-fold in DP cells (relative to P cells) and the chromosomal location. Where known, the biological role of the encoded protein from each transcript is shown. Shown in bold are the final candidate genes (*Fig 3.14*).

To determine if distinct classes of genes are de-repressed in DP cells I used the DAVID v6 programme (http://david.abcc.ncifcrf.gov/) to identify significantly enriched gene ontology (GO) categories (Dennis et al., 2003). For this analysis, GO categories containing at least 5 genes and with a threshold p-value <0.05 were selected using the mouse Ref-8 dataset as background. I used the GO biological processes (BP) database, which categorises genes according to broad biological goals, such as mitosis or reproduction, that are accomplished by ordered assemblies of molecular functions. This can be subdivided into level 1 to 5, with level 1 providing broad descriptive terms which increase in specificity but decrease in coverage (and hence statistical significance) at each successive level. I initially interrogated the broadest hierarchical GO level - biological process 1 (BP1). Interestingly, only two categories, reproduction (p = 0.0082) and developmental

processes (p = 0.032), were significantly enriched. The strong significance level of reproduction suggested that germline specific genes may be enriched among genes upregulated in DP cells. To investigate this, I interrogated the BP database at the more specific hierarchal level 3. At this level the gene-set also showed strong enrichment for germline and reproductive processes (gamete generation p = 0.0081) but also modest significance for muscle system process (p = 0.046) and response to hypoxia (p = 0.03) (*Fig 3.8a*). In contrast to genes upregulated in DP cells, no association with germline processes was found with genes downregulated in DP cells. Because germline associated genes were significantly upregulated in DP cells I investigated whether any specific tissues were enriched. Analysis of tissue specificity identified only testis (p = 0.032) and placenta/embryonic tissue (p = 0.032) as being significantly enriched among genes upregulated in DP MEFs (*Fig 3.8b*). Taken together, these data suggest that germline and testis specific genes are preferentially de-repressed in the absence of DNA methylation, consistent with previous global indications that germline restricted genes may be regulated by promoter methylation (Weber et al., 2007; Fouse et al., 2008).

To examine whether genes deregulated in DP cells were enriched in specific genomic loci, I mapped the genes to their chromosomal locations. This analysis found a significant percentage of upregulated genes were located on the X-chromosome (11.7%, expected 3.7%) (p = 0.0002, Benjamini correction p = 0.00044) (*Fig 3.8c*) as has been previously noted in hypomethylated ES cells (Fouse et al., 2008). Because these cells are male, ectopic activation of an inactive X-chromosome can be excluded as a mechanism causing enrichment of X-linked genes in DP cells. Instead, this effect is probably linked to the high proportion of testis-specific genes located on the X-chromosome (Wang et al., 2001).

To further analyse the dataset I investigated whether any specific promoter classification types were enriched among the DP upregulated genes. I used the promoter categorization system employed by Mikkelsen et al (2007) and thus divided genes into low CpG density promoters (LCP) (500bp Obs/Exp<0.4), intermediate CpG density (ICP) (500bp Obs/Exp 0.4-0.6) and high CpG density

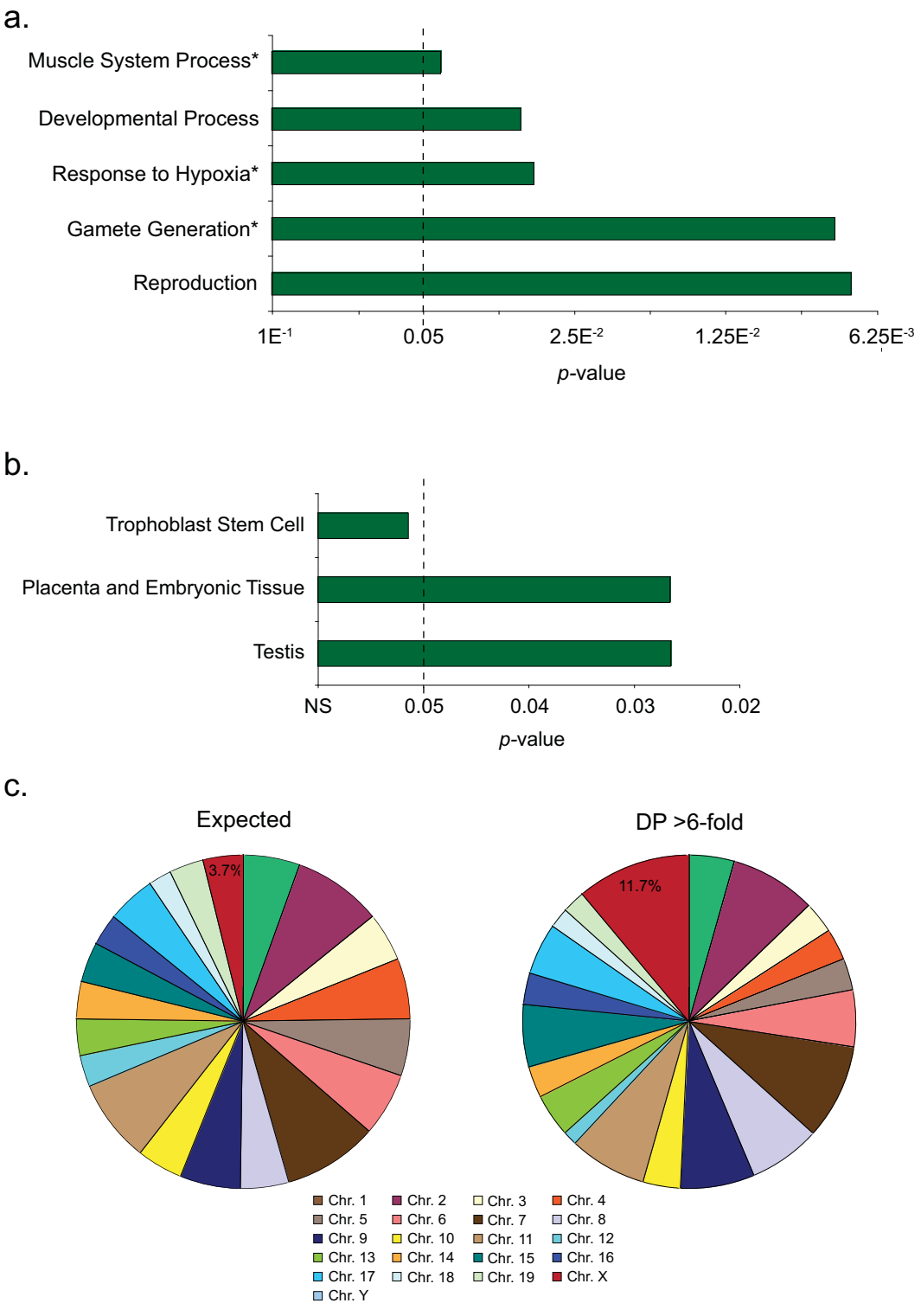


Fig 3.8. **Biological functions of genes upregulated in DP cells.** a.) Gene ontology of genes upregulated >6-fold in DP MEFs. Shown are all significant enrichment classifications at GO biological process level 1 (BP1) & biological process 3 (BP3) (asterik) using the DAVID programme. Dashed line indicates p = 0.05 significance. b.) Tissue specificity of genes de-repressed in DP cells. Only testis and placenta/embryonic tissues were significantly enriched. c.) Chromosomal location of genes upregulated in DP cells (right panel) and the expected distribution (left panel). Only genes located on the X-chromosome are significantly enriched (p = 0.00044) among genes upregulated >6-fold in DP cells.

promoters (HCP) (500bp Obs/Exp>0.6, GC>55%). It has been proposed that potential targets for regulation by DNA methylation would be associated with ICP promoters, so called weak-CpG islands. This is because the density of CpGs at LCP promoters is insufficient to promote silencing when methylated. Conversely HCPs are very rarely methylated in any tissue and therefore are unlikely be regulated by methylation. ICPs therefore represent an intermediate that has a high enough CpG density to effect transcription in *cis* yet also has the potential to acquire developmental or tissue specific DNA methylation. It would therefore be predicted that genuine direct targets of DNA methylation would be enriched in ICPs. My analysis was able to map 162 of the 221 upregulated DP transcripts to promoter classifications using the Mikkelsen dataset. These genes showed a strong deviation from the expected distribution with a 2-fold enrichment of ICP and LCP promoters. There was also a 2-fold reduction in the expected number of HCP promoters (*Fig 3.9a*). This suggests that ICP and LCP associated genes are preferentially upregulated in DP cells. The enrichment of ICP genes is consistent with the notion that this class of genes are likely candidates for regulation by DNA methylation. Conversely, the significantly reduced proportion of HCP genes probably reflects their already hypomethylated status. In contrast to genes upregulated in DP cells, downregulated genes showed no significant deviation from the expected distribution of promoter types (*Fig 3.9a*).

Interestingly, analysis of the HCP genes that *were* upregulated in DP cells demonstrated these genes were greatly enriched (10-fold) in promoters devoid of the histone modifications H3K4me3, H3K9me3 and H3K27me3 (as determined by the Mikkelsen dataset) (*Fig 3.9b*). As noted by Fouse et al (2008) HCP genes depleted in histone modifications may represent the small fraction of HCP genes that are regulated by DNA methylation. It is therefore interesting that such a striking number of HCP targets lacking histone modifications should be found among genes upregulated in DP cells. Consistent with this notion, at least two upregulated HCP genes not marked by histone modifications, *Taf7l* and *Tex13*, have been shown to be methylated in somatic tissues, in contrast to what would be expected of most HCPs (Weber et al., 2007; Meissner et al., 2008). It is noteworthy that 91% of the germline

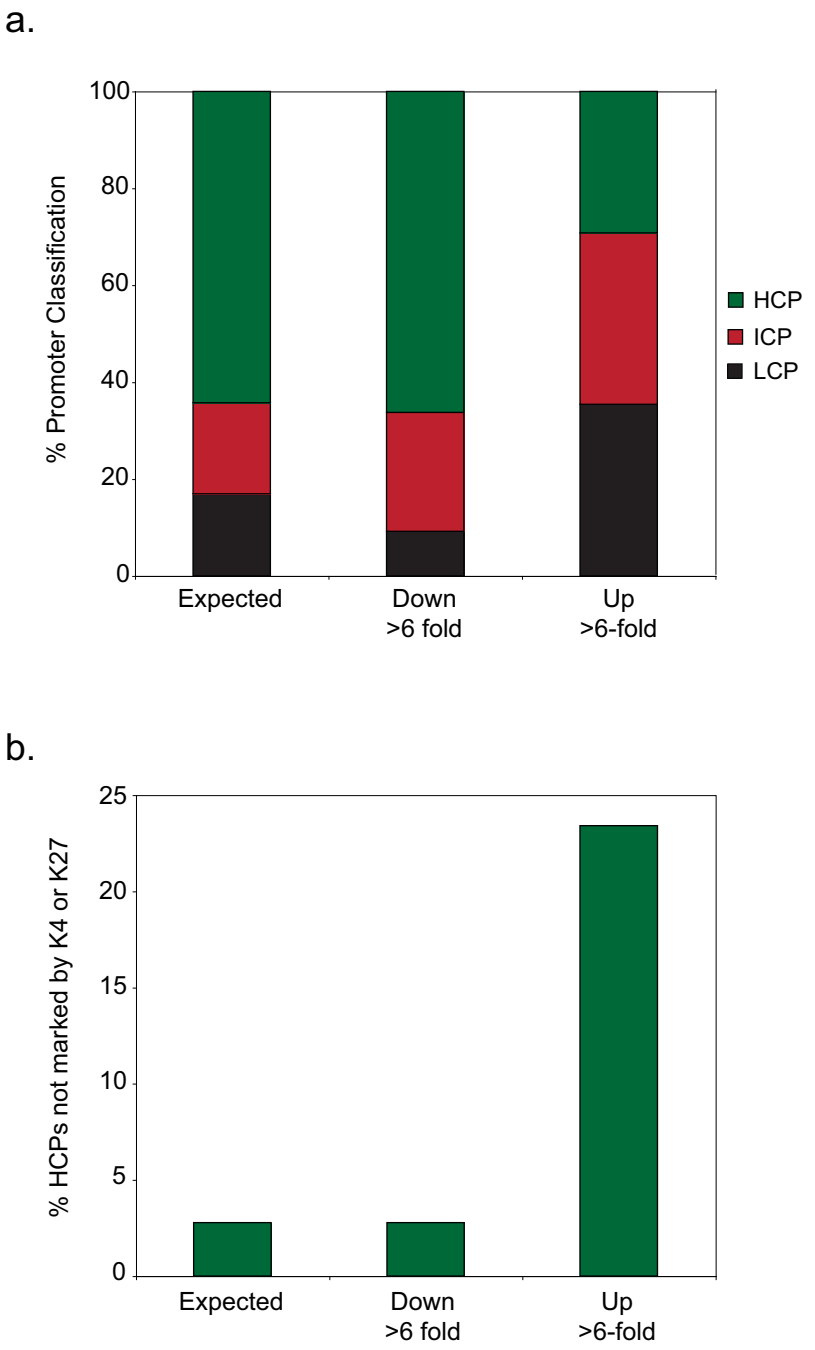


Fig 3.9. **Promoter classification of mis-expressed genes.** a.) Genes upregulated >6-fold in DP cells are strongly enriched for ICP and LCP promoters and depleted in HCP promoters. In contrast, genes downregulated >6-fold in DP cells exhibit no significant deviation from the expected distribution. b.) Analysis of the HCP genes that *are* upregulated in DP cells demonstrates they are highly enriched for loci not marked by bivalent chromatin modifications relative to the expected number. HCP genes downregulated >6-fold in DP cells show no significant difference from the expected number (p = 0.79).

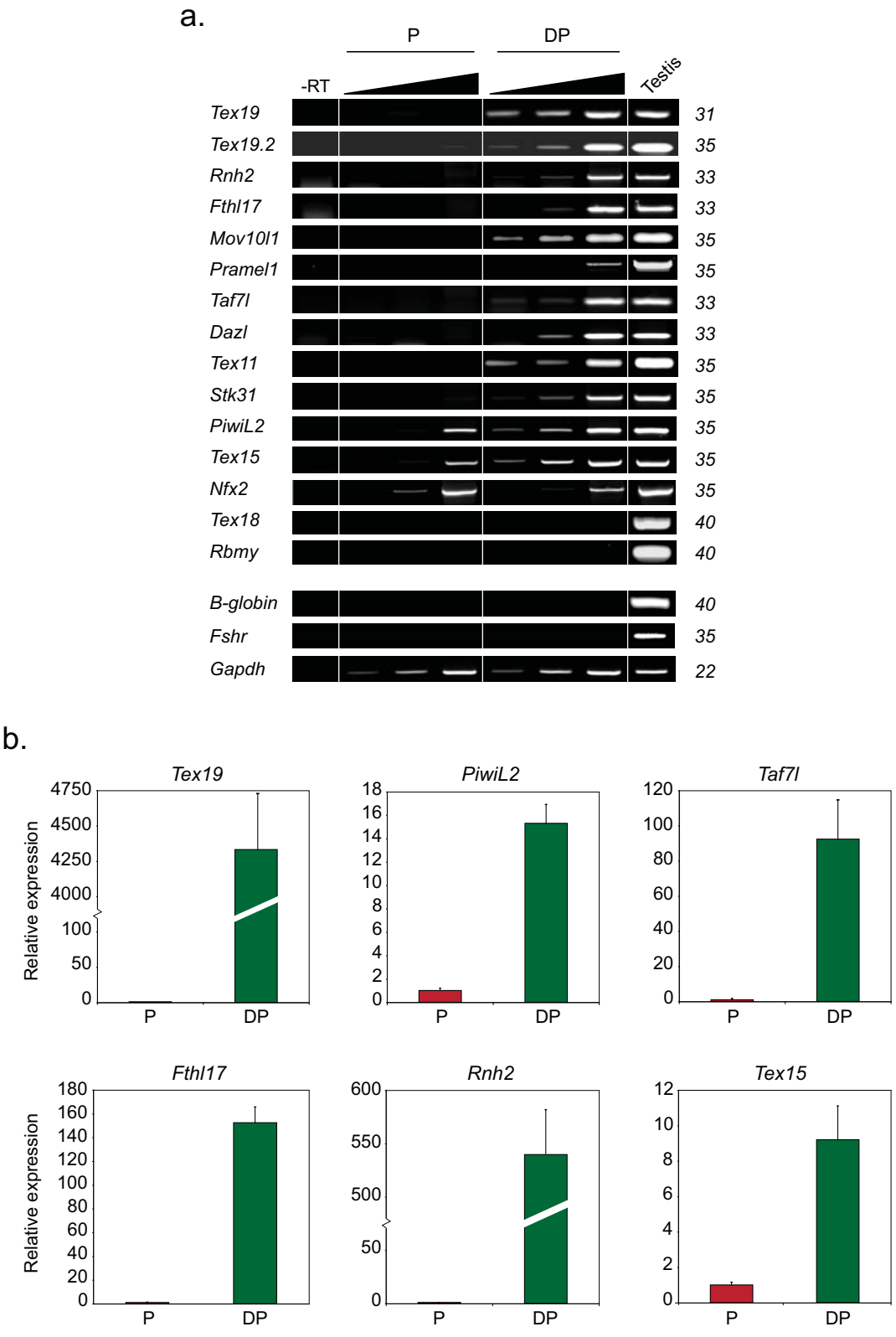


Fig 3.10. **Germ-line specific genes are de-repressed in DP MEFs.** a.) Semi-quantitative RT-PCR of a cohort of germ-line specific genes (Wang et al., 2001). Analysis was performed over increasing cDNA dilutions of 1/50, 1/20 and 1x (black curve). Number of cycles is shown on the right-hand side. *B-globin* and *Fshr* are negative controls. *Fshr* is only expressed in the somatic cells of the testis. *Gapdh* is a loading control b.) Quantitative RT-PCR analysis of selected germline specific genes normalised to *Gapdh*. Shown is fold change relative to P MEFs, which are set to 1. Error bars are S.E.M.

specific genes upregulated in DP cells fall into the ICP or HCP category. This is consistent with germline genes being direct targets of DNA methylation rather than indirect hits as would be predicted of LCP genes.

Taken together, these analyses suggest that germline specific genes may be targets for regulation by DNA methylation. To further investigate the role of DNA methylation at germline associated genes I conducted a semi-quantitative RT-PCR analysis based on a set of testis-specific genes reported by Wang et al (2001). This set of genes was selected based only on their reported testis-specific expression pattern and irrespective of the array analysis. Remarkably, RT-PCR showed that 12 of the 15 testis-specific genes tested were strongly expressed in DP cells but at undetectable or significantly reduced levels in P cells (*Fig 3.10a*). This suggests that de-repression of germline associated genes in DP cells is a general phenomenon. I over-cycled each reaction (usually 35 cycles) to demonstrate that in most cases (80%) a transcript could not be detected in P cells, as would be expected for the germline-specific nature of these genes. Quantitative RT-PCR of six genes confirmed they were de-repressed between 8-fold and >4000-fold in DP cells (*Fig 3.10b*). In contrast, I was unable to detect any transcripts for ES cell specific genes in either P cells or DP cells (*Fig 3.7b*). These data support the conclusion of the gene ontology analysis that germline specific genes are highly enriched among genes de-repressed in DP cells. To refine the 221 genome-wide candidate genes generated here to a more specific and testable set, I resolved to profile P cells treated with 5-aza dC and cross-reference de-repressed genes with those candidates identified generated here.

3.6 Germline specific & immune response genes are de-repressed in 5-aza dC treated P cells

Candidate target genes regulated by promoter methylation would be expected to activate in response to the demethylating agent 5-aza dC. To test this, I treated P cells with 1μM 5-aza dC (P-aza) for 72hrs and profiled global gene expression using Illumina mouse Ref-8 BeadChips. At this concentration of 5-aza dC, p53-null cells lose ~60% of global methylation but retain viability and continue to proliferate

(Nieto et al., 2004). The array analysis identified 308 genes (1.7%) upregulated greater than 6-fold in P-aza cells. To identify gene ontology categories significantly enriched in this gene set I interrogated the BP1 database using DAVID v6. This analysis showed that, similar to hypomethylated DP cells, reproduction ($p = 0.00045$) and developmental processes ($p = 0.00022$) were significantly enriched in P-aza cells. Interestingly, at the more specific hierarchical level BP3, germline processes were still significantly enriched (gamete generation $p = 0.0008$) but several immune response categories were also moderately enriched (innate immune response $p = 0.045$, response to wounding $p = 0.0044$, leukocyte activation $p = 0.0067$) (*Fig 3.11a*). This analysis is consistent with gene ontologies from DP cells and supports a direct role for DNA methylation in regulating germline specific genes. In contrast, because immune associated genes are not enriched in DP cells, this class of genes may be indirectly activated in P-aza cells, probably as a result of the reported cytotoxic effects of 5-aza dC (Oka et al., 2005). Consistent with this, more genes were found to be upregulated in P-aza cells (308) than DP cells (221) suggesting 5-aza dC has at least some methyl-independent effects on gene expression here.

The global analysis carried out on P-aza MEFs showed a strong enrichment of germline restricted genes. To examine this further, I investigated the expression levels of the set of testis-specific genes shown in *Fig 3.10* after 5-aza dC treatment by RT-PCR (Wang et al., 2001). I determined the effect of 1μM 5-aza dC on both P cells (P-aza) and pMEFs (pMEF-aza) treated for 48hrs or 96hrs and also, the additive effect of 5-aza dC treatment (96hrs) with the histone deacetylase inhibitor TSA (*Fig 3.11b*). All the testis-specific genes identified as being de-repressed in DP cells were also expressed in P-aza cells after 48hrs and 96hrs treatment. However, some genes activated in P-aza and DP cells were not reciprocally activated in pMEFs treated with 5-aza dC. For example, *Rnh2* and *Stk31* transcripts are clearly detectable in DP and P-aza cells but not in pMEF-aza cells (*Fig 3.11b*). This could indicate that p53 plays a direct or 'priming' role to sensitise cells to demethylation. Alternatively pMEFs may be less sensitive to 5-aza dC than P or DP cells and may require a higher dose to induce demethylation, consistent with reports that suggest cell responses to 5-aza dC vary widely (Qin et al., 2009). It is also possible that genes not activated in

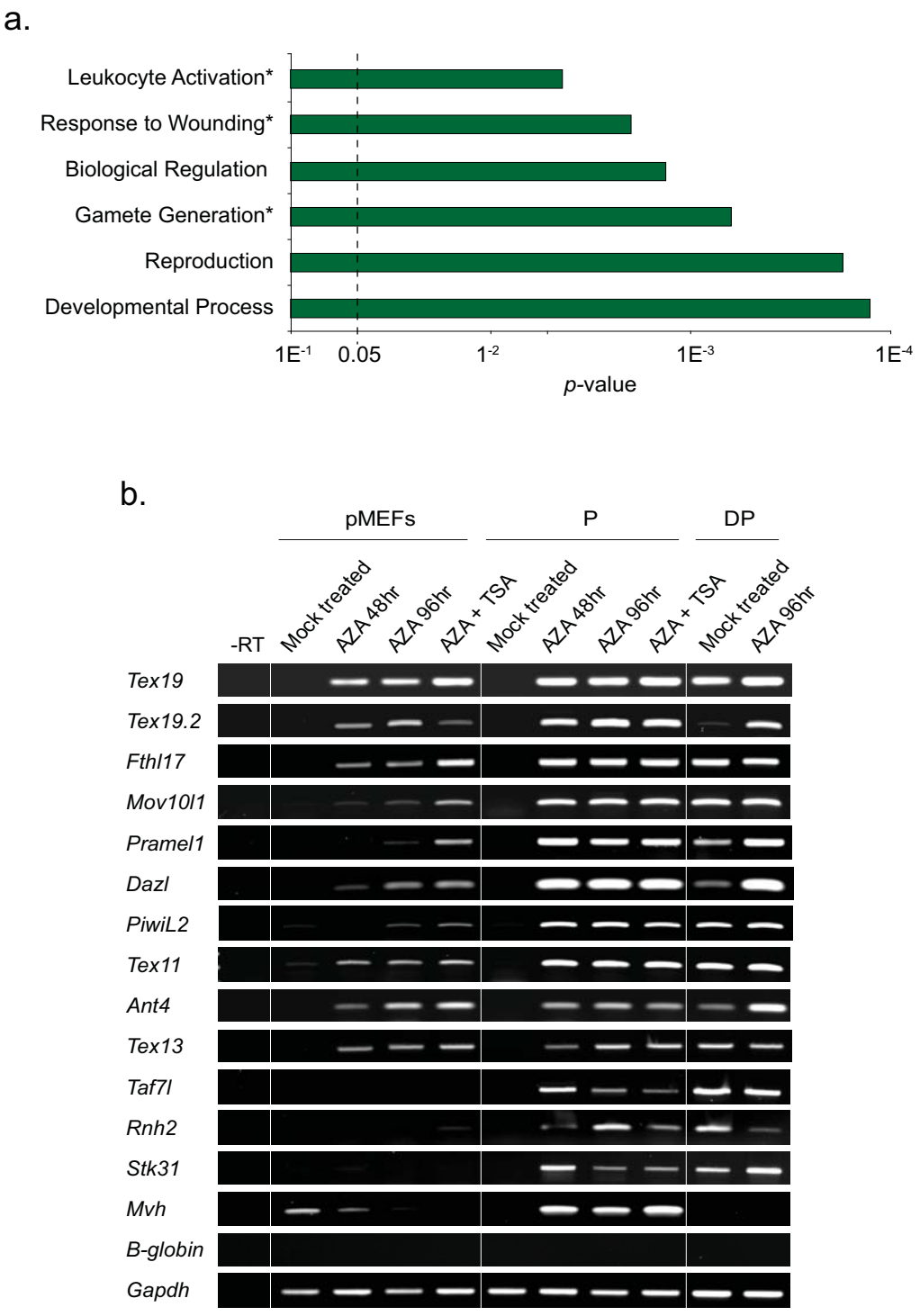


Fig 3.11. **Analysis of 5-aza dC treated P cells.** a.) Gene ontology of genes upregulated >6-fold in P-aza cells at BP1 and BP3 (asteriks). Reproduction (p = 0.00042) and developmental process (p = 0.00021) were most significantly enriched. b.) RT-PCR of pMEFs or P cells treated with 1uM 5-aza dC for 48hrs or 96hrs +/- TSA. All genes de-repressed in DP cells (far right columns) were also activated by 5-aza dC in P cells (P-aza) and the majority were activated in pMEFs (pMEF-aza). *Gapdh* is a loading control and *B-globin* demonstrates there is no general activation of tissue-specific transcripts.

pMEF-aza cells are not direct targets for regulation by DNA methylation or may require specific transcription factors not present in pMEFs but present in P and DP cells. Interestingly, *Mvh* was not activated in DP or pMEF-aza cells but was in P-aza cells (*Fig 3.11b*). This apparently inconsistent response indicates that at best *Mvh* is indirectly regulated by DNA methylation changes. This is intriguing because *Mvh* is one of three genes reported to be temporally regulated by promoter methylation during PGC development (Maatouk et al., 2006). Because this conclusion was reached by the authors based on correlative evidence, this example highlights the need for comprehensive cause and effect studies on putative methylation-dependent genes.

In my analysis, the addition of TSA did not significantly enhance de-repression of any of the germline-specific genes examined, suggesting they are primarily responsive to demethylation, and histone deacetylation is not additive with this effect. To confirm TSA cannot de-repress germline-specific genes independently of 5-aza dC, I treated P cells and pMEFs with TSA only (*Fig. 3.12a*). Here, no transcripts were detected suggesting these genes fall into the category of genes previously described by Schuebel et al (2007) and Lande-Diner et al (2006) that are activated by demethylation without the requirement for forced deacetylation and alterations to chromatin structure. I conclude that the majority (77%) of germline specific genes examined here are de-repressed in all cell-types examined by a relatively low concentration (1μM) of 5-aza dC but not TSA, indicating they could be targets for direct regulation by DNA methylation. This, combined with the global analysis of P-aza cells and DP cells strongly supports a role of DNA methylation in regulating germ cell-specific genes.

To generate a refined list of candidate genes I cross-referenced the genome-wide transcripts de-repressed in both P-aza cells and DP cells. Eighty seven genes were found to overlap between the global experiments whereas 134 and 221 genes were upregulated in only DP cells or P-aza cells, respectively (*Fig 3.12b*). The relatively small degree of crossover (16.5%) suggests that de-repression of many genes may be a consequence of stochastic or indirect effects. However, the genes present in both

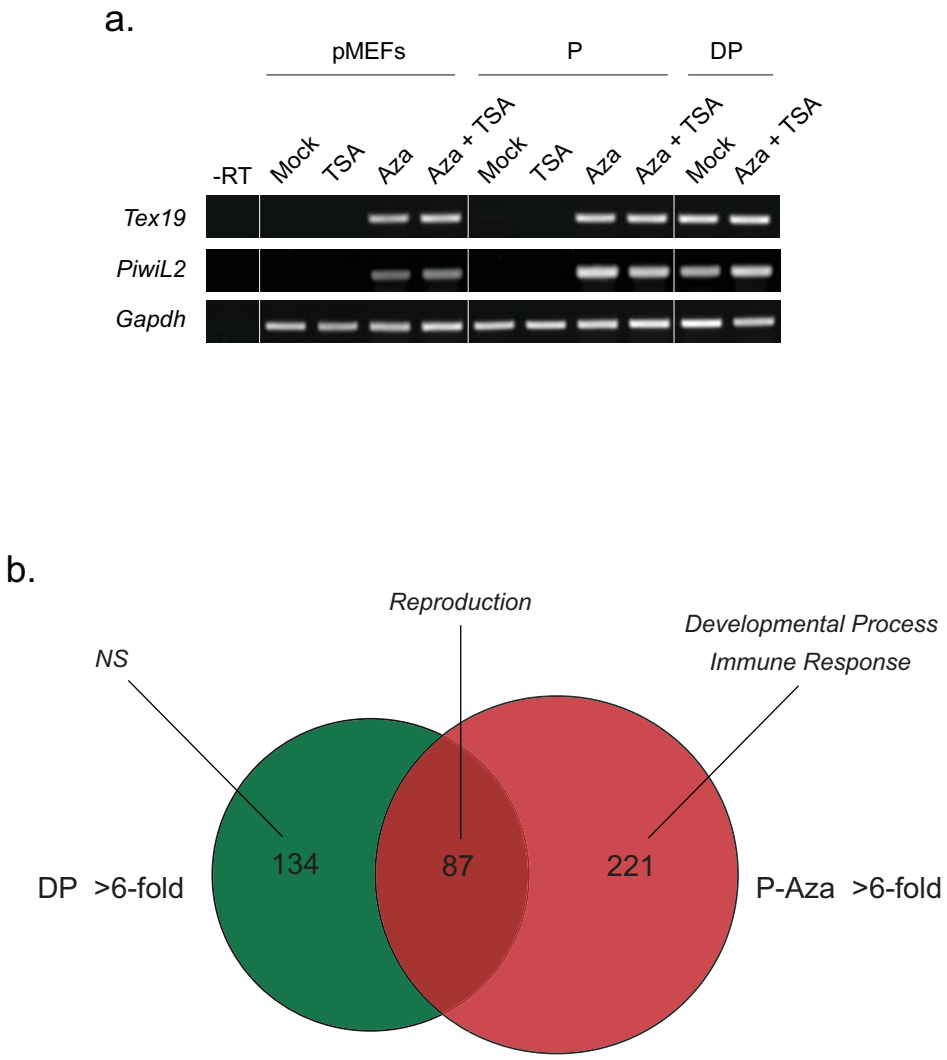


Fig 3.12. **Cross-referencing genes upregulated in DP and P-aza cells reveals enrichment for germline GO categories.** a.) The candidate genes *Tex19* and *PiwiL2* are not activated by TSA but are de-repressed in response to 5-aza dC in pMEFs and P cells. b.) Venn diagram cross-referencing genes upregulated in DP cells and P-aza cells shows 87 overlapping candidates. Overlapping genes are strongly enriched in germ-line associated genes ($p = 0.0005$) at BP1. The remaining genes de-repressed in DP cells showed no significant enrichment (NS) whereas the remaining genes only de-repressed in P-aza cells were enriched for immune response and developmental classifications.

datasets (overlap) would be predicted to be enriched in targets directly regulated by DNA methylation. Strikingly, BP1 GO analysis showed that only the overlapping gene set was significantly enriched for only germline associated genes (reproduction $p = 0.0005$). In contrast, the remaining DP and 5-aza dC datasets, lacking the overlapping genes, show no significant enrichments (DP) or enrichments for developmental and apoptotic GO classifications (P-aza cells). This suggests that the genes responsible for the germline gene ontology enrichment in the DP and P-aza datasets are the same genes that overlap between the two gene lists. This strengthens the notion that germline genes are strong candidates for regulation by promoter CpG methylation. Furthermore, this analysis has identified a refined list of 87 germline-enriched candidate genes, which are upregulated in both genetically and pharmacologically hypomethylated cells.

3.7 Epigenetic memory of germline specific genes is mediated by DNA methylation

To reduce the number of candidate genes to a testable figure and to enrich with genuine methylation-dependent targets, I resolved to undertake a third experimental approach to identify genes which rely on promoter CpG methylation for epigenetic memory. In this approach candidate targets regulated exclusively by DNA methylation would be predicted to activate in response to 5-aza dC and crucially, remain expressed indefinitely following 5-aza dC removal. Here, the rationale predicts that genes that are indirectly activated or regulated by additional epigenetic mechanisms would re-impose repression during a recovery period. For example, indirect targets activated as a consequence of alterations to transcription factor networks would return to their original expression state coincident with cellular recovery to normal physiology. Conversely, candidate loci where DNA methylation may have a maintenance or secondary role in epigenetic silencing would be predicted to re-target repression through alternative epigenetic mechanisms, such as H3K27me3, H3K9me2 or antisense RNAs (Vire et al., 2006; Lachner et al., 2001; Feladma et al., 2006; Chotalia et al., 2009). In contrast genes that critically rely on only DNA methylation for silencing in somatic cells would remain expressed even

after removal of 5-aza dC as they would have lost their epigenetic memory (Feng et al., 2006).

To test for candidate loci that rely on promoter methylation for epigenetic memory I treated P cells and mouse NIH/3T3 cells with 1μM 5-aza dC for 72hrs, followed by a 14 day recovery period under normal culture conditions. Following 5 days of recovery P cells exhibited 50-70% cell death, probably as a result of chromosomal aberrations and p53-independent apoptosis (Nieto et al., 2003). In contrast NIH/3T3 cells appeared morphologically normal and proliferated as expected with only 10-30% cell death. For this reason I chose to use NIH/3T3 cells in this analysis. I profiled the global expression levels in NIH/3T3 cells prior to treatment (3T3-control), immediately following treatment (3T3-aza) and after 14 days recovery (3T3-recovery) using Illumina mouse Ref-6 BeadChips. This additionally allowed me to compare 5-aza dC induced genes from P cells and NIH/3T3 cells. It is noteworthy that the Ref-6 arrays used to profile 3T3 cells here contain 30,775 transcripts compared to 18,098 transcripts for the Ref-8 arrays used for the previous analysis. Therefore in direct comparisons, percentages of the arrayed dataset are also given.

This analysis identified 423 genes (1.4%) upregulated >6-fold in 3T3-aza cells. This compares similarly to the 1.7% observed in P-aza cells. In contrast, after 14 days recovery, only 76 genes (0.25%) were significantly upregulated in the 3T3-recovery cells. Of the genes that were upregulated after 14 days recovery, 54 (71.1%) were also upregulated >6-fold in the 3T3-aza geneset (overlap) and therefore were turned on by 5-aza dC and crucially, remained on (*Fig 3.13a*). These 54 overlapping genes represent those which have putatively lost their epigenetic memory and cannot re-impose silencing following demethylation. The remaining 22 genes present exclusively in the 3T3-recovery geneset were largely (n=18) just below (3-6 fold) the >6-fold threshold in the 3T3-aza list, indicating that technical considerations (the threshold value) rather than late-onset upregulation *per se*, excluded them from the overlapping geneset. Interestingly, BP1 GO analysis of the overlapping genes, which have lost transcriptional memory, showed enrichment for only the germline

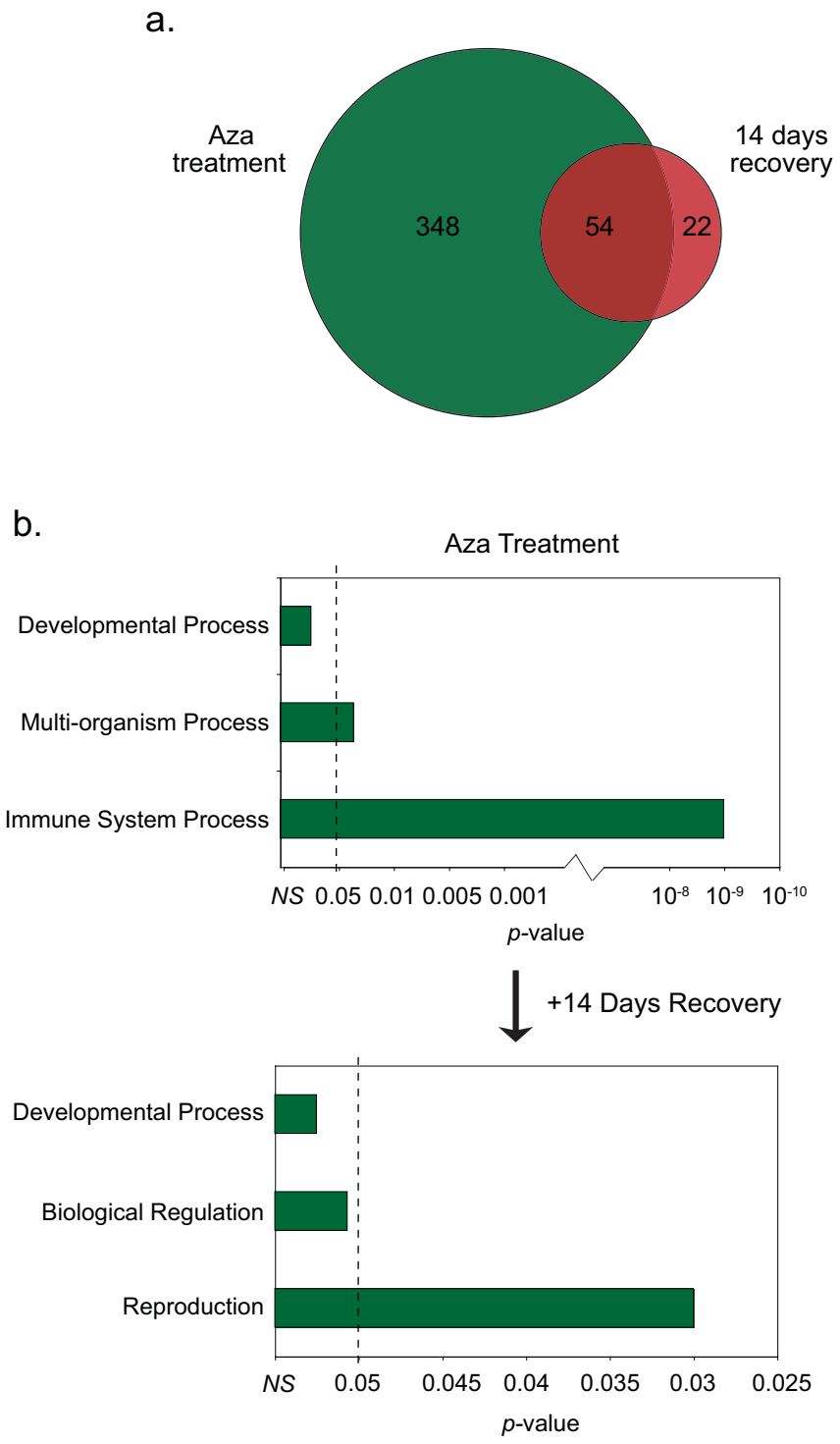


***Fig 3.13*. Germline specific genes preferentially rely on DNA methylation for epigenetic memory** a.) Venn diagram showing the overlap between genes de-repressed by 5-aza dC treatment in 3T3 cells (left circle) and following a 14 day recovery period in the absence of 5-aza dC (right circle). Overlapping genes were significantly enriched in germline specific genes ($p = 0.03$). Dashed line indicates a significance threshold of $p = 0.05$. b.) Gene ontology analysis at BP1 of 3T3-Aza treated cells shows 5-aza dC treatment preferentially activates immune response ($p = 1.1^{-9}$) and multi-organism process ($p = 0.042$), whereas after 14 days recovery only reproduction is significantly enriched ($p = 0.03$) among genes still upregulated >6-fold (overlap). *NS* = not significant

associated category, reproduction (p = 0.03). In contrast, the 3T3-aza geneset showed significant enrichment for immune system processes ($p = 1.1^{-9}$) and multi-organism process (p = 0.0043) but not reproduction (*Fig 3.13b*). This data strongly suggests that germline genes are activated in globally hypomethylated cells and crucially, in contrast to other gene categories, are unable to re-impose gene silencing. This suggests that in addition to being silenced by CpG methylation, germline specific genes critically rely on DNA methylation for epigenetic memory. To validate this, I monitored the expression of germline genes during 5-aza dC recovery in both P cells and pMEFs by RT-PCR (*Fig 3.14a &* b). In agreement with the array analysis, the germline associated genes remained expressed following 5-aza dC treatment whereas control genes rapidly re-imposed gene silencing. I conclude that germline genes likely represent a distinct subclass of genes that depend on DNA methylation for repression but uniquely, this dependence confers an epigenetic memory that cannot be regained once removed in somatic cells.

3.8 Candidate genes

To derive a final set of candidate genes for further analysis I cross-referenced data from all three experiments (*Section 3.5, 3.6 & 3.7*) (*Fig 3.14c*). This analysis identified 14 single-copy genes and additionally, IAP elements, as upregulated in each of these stringent experiments (*Table 3.2*). Interestingly, all 14 genes exhibited strong tissue-specificity. Of the fourteen genes, nine are primarily expressed in the testis with the remaining expressed in liver, kidney, blood and the pituitary gland according to the BioGPS and UniGene databases. Furthermore, of the mappable genes, 10 out of 13 are associated with ICP or HCP promoters. Interestingly, all eight mappable testis-specific genes are on the classification border between ICP and HCP promoters i.e. weak CpG island genes, which is precisely what would be expected of genes potentially regulated by DNA methylation (all have CpG Obs/Exp 0.57-0.61 over >400bp) (*Fig 4.1*). Indeed, the only gene from this gene list, *Tex13*, which was included in the global bisulphite sequencing analysis by Meissner and colleagues (2007) was strongly hypermethylated in somatic cells despite being classified as a HCP gene (see *Table 3.2*.). In contrast, 3 out of 5 of the remaining non germline

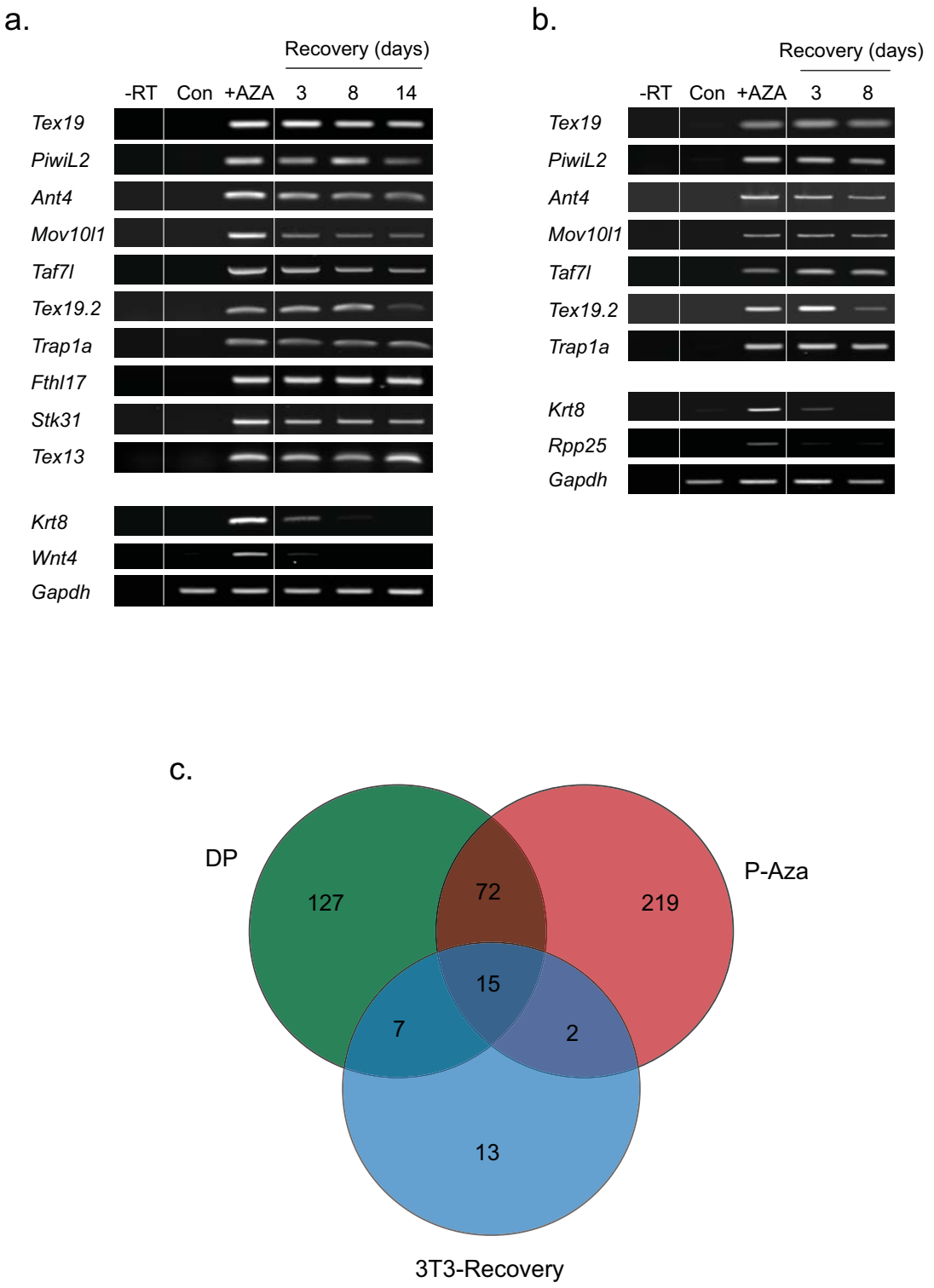


Fig 3.14. **Identification of final candidate genes.** a & b.) RT-PCR of germ-line genes in a.) P cells and b.) pMEFs after 5-aza dC treatment and at days 3, 8 and 14 following recovery after 5-aza dC withdrawl. In contrast to germ-line genes, the controls *Krt8, Rpp25* and *Wnt4* re-impose silencing during recovery from 5-aza dC. c.) Venn diagram cross-referencing all three experiments (*Fig 3.4*) to identify the final candidate genes. This analysis suggests 15 transcripts (14 single-copy) are upregulated >6-fold in all three experiments and therefore satisfy the stringent criteria to be considered candidate genes here. Notably, additional gene probes present in the 3T3 recovery experiment have been excluded. Not to scale.

genes are associated with LCP promoters. This suggests that these genes may be indirect hits as a result of other epigenetic or transcription factor changes. Alternatively, it is possible DNA methylation could be regulating their expression through distal CpG rich elements. For example, a CpG island is present 2kb upstream of the *Gstp2* promoter. However, the prevalent gene ontology conclusion reporting enrichment of germline specific genes, considered with the promoter classifications, suggest that the nine testis-specific genes identified here represent the most promising targets for further research.

Gene	Genomic location	Tissue specificity	Promoter type	Methylated
Tex19	11	Testis	HCP	ND
PiwiL2	14	Testis	ICP	ND
Rps4y2	6	Testis	HCP	ND
Rhox5	X	Testis	Unknown	ND
MageK1	X	Testis	ICP	ND
Tex13	X	Testis	HCP	Yes
Mov10l1	15	Testis	ICP	ND
Rhox4d	X	Testis	ICP	ND
Tex19.2	11	Testis	ICP	ND
Gpr97	8	Blood	LCP	ND
Slc47a1	11	Kidney	HCP	ND
Akr1c12	13	Liver	LCP	ND
Prl	13	Pituitary Gland	LCP	ND
Gstp2	19	Liver	ICP	ND
IAP element	N/A	(Testis)	ICP	Yes

***Table 3.2.* Methylation-dependent candidate genes.** Candidate genes were derived by cross-referencing genes de-repressed in the three experimental approaches used here (*Fig 3.4 & 3.14c*). N/A: Not applicable, ND: Not determined. Nb. IAP is expressed preferentially but not exclusively in the testis.

In summary, my approach to progressively identify potential candidates for regulation by DNA methylation has yielded a set of 14 single copy genes. Of these, 9 are testis-specific and have the promoter characteristics expected of methylation-dependent targets. Further analysis (*Chapter 4*) will investigate whether these genes are causally regulated primarily by CpG methylation *in vivo.*

3.9 Germline specific miRNAs and piRNAs are not de-regulated in DP cells

This chapter identified a significant enrichment of germline specific genes de-repressed in the absence of DNA methylation. I decided to examine whether this germline specificity could be extended to other varieties of transcript such as testis specific microRNAs (miRNA) and piwi-interacting RNAs (piRNAs). piRNAs are short RNAs typically 26-32bp in length specifically expressed during spermatogenesis where they function to target a silenced chromatin state to transposable elements and some genes (Klattenhoff & Theurkauf, 2009). To investigate whether piRNAs were globally mis-expressed in DP cells we γ-^{32}P labelled total RNA from wt testis, P cells and DP cells (collaboration with *Dr A. Ruzov,* MRC HGU). Separation of RNA through polyacrylamide clearly showed strong bands relating to ~30bp in wt testis but not in either P or DP cells (*Fig 3.15a*). This suggests that DNA methylation is not crucially required for regulating global piRNA expression in somatic cells. Alternatively, DNA methylation may have a role in piRNA regulation but components of the piRNA biogenesis machinery are absent in DP cells preventing piRNA production.

To examine the effect of genome-wide hypomethylation on testis specific miRNA expression I performed miRNA specific qPCR. Here I was unable to detect significant expression of any of the six testis-specific miRNAs tested in pMEFs, P or DP cells compared to wild-type testis (*Fig 3.15b*) (Ro et al., 2007). To extend this analysis I collaborated with *Dr E. Wade* (University of Leicester) to profile genome-wide miRNA expression in hypomethylated DP and control P cells. This array analysis failed to identify any miRNA's significantly mis-expressed (>2-fold) between the cell lines and only limited miRNA with >1.5 fold changes (*Appendix 3*). This work indicates that while DNA methylation has a potentially important role in regulating germline specific genes, it does not appear to significantly regulate, testis-specific miRNAs and piRNAs or indeed the global miRNAs tested here in somatic cells.

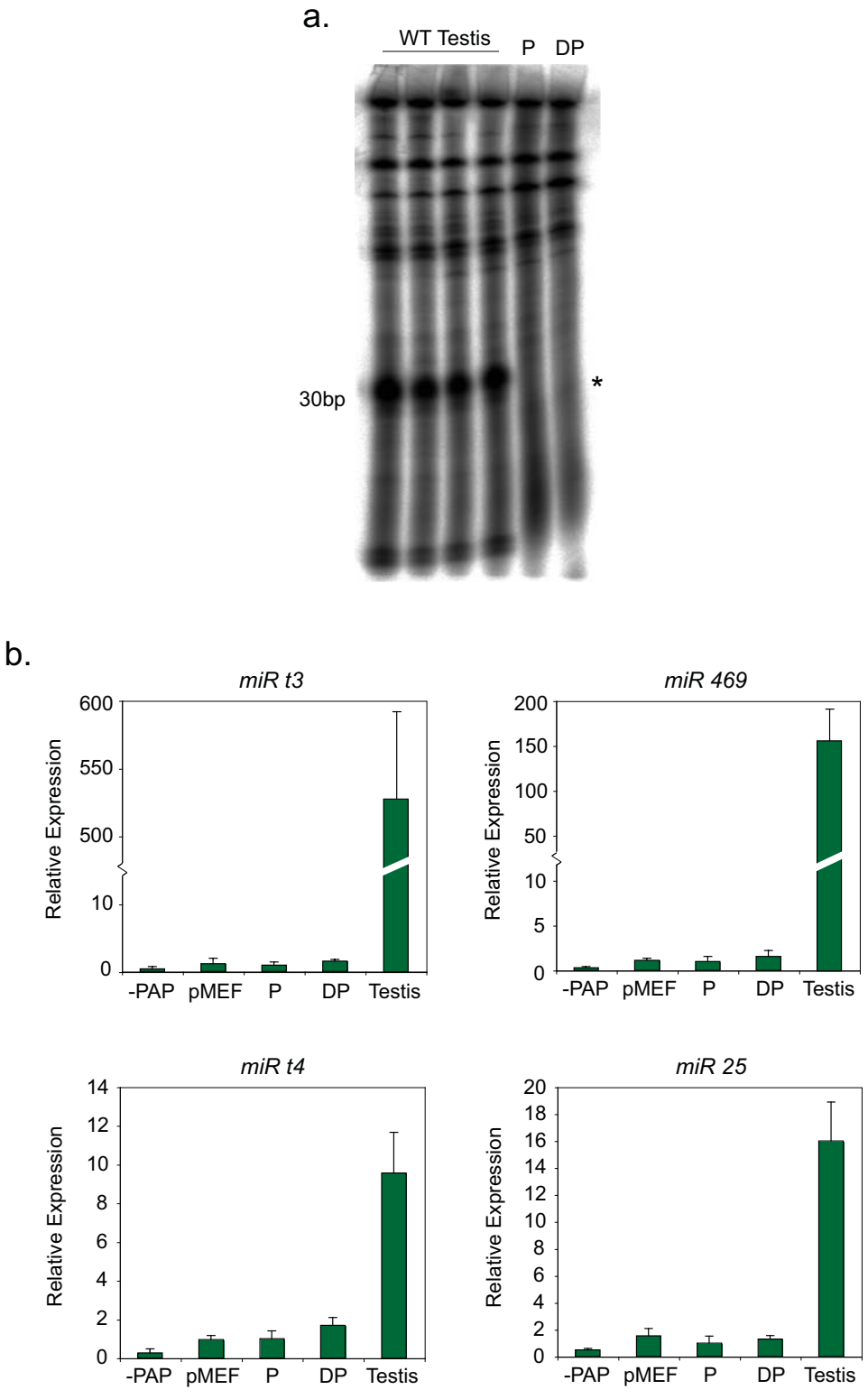


***Fig 3.15*. piRNAs and testis-specific microRNAs are not de-repressed in DP cells.** a.) Radiolabelled total RNA from four WT testis, P cells and DP cells was size separated through a poly-acrylamide gel. Asterik shows the expected migration of piRNAs at ~31bp. Total piRNAs were not detectable in either P cells or DP cells relative to the testis positive control b.) qRT-PCR of 4 testis-specific miRNAs in pMEFS, P cells, DP cells and testis positive control. Shown is expression relative to P cells, which is set to 1. No signifcant miRNA expression was detected relative to testis. -PAP = - Poly-A polymerase negative control.

3.10 Discussion

In order to ascertain whether DNA methylation has a primary role in regulating developmental or tissue-specific gene expression, it is first necessary to identify candidate loci. Previous studies to identify candidate genes have used either a single functional assay (i.e. WT vs hypomethylated cell expression array) (Jackson-Grusby et al., 2001; Fouse et al., 2008) or differential methylation analysis (Suzuki et al., 2007; Weber et al., 2007; Rodic et al., 2005). These methods have the drawback of generating many indirect hits and false-positives or targets where it is not clear whether differential methylation is a consequence or a cause of transcriptional activity, respectively. Furthermore, often these analyses are not validated by downstream cause and effect experiments. Thus, I set out to identify a set of *bona fide* candidate genes that could be used in further causal analysis of the role of DNA methylation. I used a novel functional approach to progressively identify candidate loci by selecting genes that fulfil each of three independent experimental predictions (*Fig 3.4*). By cross-referencing the three experimental datasets I minimized false-positives and indirect hits. Additionally, the highly stringent 6-fold expression threshold enriched for genuine targets at each experimental stage. This stringent multi-tiered approach has led to the identification of fourteen single-copy candidate genes. These genes are activated when DNA methylation is removed either genetically or pharmacologically. Crucially, they also appear to rely on methylation to retain an epigenetic memory of their lineage specific transcriptional status.

An intriguing and novel aspect of my approach was the final assay to identify genes that lose their epigenetic memory when DNA methylation is transiently removed by 5-aza dC treatment. The majority of de-repressed genes (71.1%) were able to re-impose gene silencing during a recovery period in the absence of 5-aza dC. However 54 genes remained highly transcriptionally active throughout this recovery period. These genes represent targets that have lost their epigenetic memory and therefore cannot re-acquire their original lineage-specific expression state. Unlike genes able to re-impose silencing, it would be predicted that these genes critically rely on DNA methylation to direct transcriptional memory. Interestingly, targets that lost their

epigenetic memory in this assay were highly enriched for germline specific genes. Indeed, in each experimental filter germline genes were strongly enriched among de-repressed targets and nine of the final fourteen candidate genes are testis-specific. Taken together, the experimental data generated here strongly suggests that at least a subset of germline specific genes could be directly and primarily regulated by DNA methylation. In support, the candidate genes identified here did not overlap with genes that rely on the H3K9 methyltransferase G9a to target and/or maintain promoter DNA methylation (Epsztejn-Litman et al., 2008). This is consistent with the epigenetic memory of my candidate genes being mediated by CpG methylation *per se* and not maintained or re-targeted by G9a and/or H3K9me2 as has been demonstrated for several other loci (Tachibana et al., 2008; Feldman et al., 2006).

Using my stringent experimental approach, several genes previously reported to be putatively regulated by CpG methylation were excluded from the candidate list. For example, previous studies have reported that *Pgk2* and *Mvh* directly rely on promoter methylation for their *in vivo* expression pattern (Ariel et al., 1991; Maatouk et al., 2006). However, in the present analysis these genes were not de-repressed in DP cells and showed a cell-type specific response to 5-aza dC treatment. Similarly, *Pdha2* and *Mael* were not de-repressed in any assays performed here (Iannello et al., 2000; Xiao et al., 2009). These data support a potential secondary role for DNA methylation at these loci but are not compatible with promoter methylation being the primary system for regulating their expression, as has been suggested. Moreover, a recent study elegantly demonstrated that *Elf5* relied on promoter methylation in epiblast cells to maintain embryonic lineage restriction (Ng et al., 2008). However, the results generated here indicate that *Elf5* was not de-repressed in DP cells of P-aza cells. This suggests that while promoter methylation of *Elf5* is crucially important during early embryonic stages (Ng et al., 2008), other factors contribute to silencing of *Elf5* in somatic cells and CpG methylation therefore cannot be the primary system at this locus in somatic cells. The fact that several genes previously reported to be regulated by promoter methylation did not pass the stringent experimental filters used here, highlights the requirement for cause and effect studies to validate the precise role of DNA methylation at each locus.

While several genes previously reported to be regulated by CpG methylation clearly failed the parameters set here, some putative methylation-dependent genes were only excluded because they failed a single criterion. It is possible these genes represent genuine methylation-dependent loci, but were excluded here as a consequence of the necessity to impose strict thresholds (>6-fold) to enrich for the most promising candidate genes. One example is *Dazl,* which was significantly upregulated in DP cells and P-aza cells but which partially re-imposed silencing below the threshold during recovery from 5-aza dC treatment (~10-fold activation by 5aza dC to ~4-fold activation after recovery). Thus, the candidate genes generated here are likely not an exhaustive list of potential methylation-dependent targets but rather the most promising loci. Indeed, it is noteworthy here that the Illumina BeadChiP probe for a highly expected methylation-dependent candidate, *Ant4,* was poorly designed and therefore this gene does not feature in my analysis (Rodic et al., 2005).

In summary, my approach was designed to identify candidate targets where DNA methylation is the crucial and upstream system for imposing gene silencing and maintaining transcriptional memory. This analysis identified 14 single-copy targets, which are greatly enriched in germline specific genes. To further investigate and confirm that promoter CpG methylation is the *causal* and primary *in vivo* regulator of these genes, I elected to examine two candidates, *Tex19* and *PiwiL2* in detail, as paradigms for germline-specific genes regulated by DNA methylation.

Chapter 4

Promoter CpG methylation developmentally regulates *Tex19* and *Piwil2*

4.1 Introduction

Whether promoter DNA methylation patterns are the cause, a consequence or a contributor to gene expression patterns is a question of great debate. The mounting body of evidence suggests that the precise function of promoter CpG methylation in regulating gene expression is locus specific, depending on the contribution from several epigenetic and trans-acting factors. For example at the *Oct3/4* promoter, CpG methylation is sufficient to maintain repression but is neither necessary nor the upstream initiating signal for gene silencing (Feldman et al., 2006). Instead CpG methylation at this locus co-operates with other epigenetic systems and trans-acting factors to contribute to a complex regulatory system (Gu et al., 2005; Cedar & Bergman, 2009). At the *Magea2* locus both H3K9me2 and promoter methylation are independently sufficient to maintain gene silencing but neither is crucially required *per se* (Tachibana et al., 2008). Here, both epigenetic marks can promote silencing and thus each functions as an additional layer of repression to reinforce the epigenetic expression state. Several silenced *Hox* gene promoters also accumulate CpG methylation but here, DNA modification acts downstream of polycomb mediated repression (Reynolds et al., 2006; Vire et al., 2006; Fouse et al., 2008). At these loci, promoter methylation may contribute to maintaining gene repression but is generally considered to be a consequence of transcriptional silencing.

In addition to epigenetic regulation, many genes require tissue-specific transcription factors to drive expression. This can add an additional layer of regulation such that permissive epigenetic marks and appropriate transcription factors must both be

present. For example, in somatic cells demethylation of the *Elf5* promoter is necessary but not sufficient for gene activation due to the absence of trophoblast specific transcription factors (Ng et al., 2008). Indeed, the testis specific *Rhox6* and *Rhox9* genes, which have been proposed to be regulated by DNA methylation, are not activated in hypomethylated *Dnmt1*-null ES cells but are aberrantly expressed when these mutant ES cells are differentiated (Oda et al., 2006). This indicates that either additional mechanisms of epigenetic silencing aside from DNA methylation operate at *Rhox6* and *Rhox9* in ES cells or that there is a change in the availability of a limiting transcription factor during differentiation. Thus, the emerging picture is that promoter methylation can contribute to maintenance of an expression state but is largely a part of complex gene-specific repression systems that include diverse histone modifications, histone variants and transcription factor networks (Hemberger et al., 2009; Reik, 2007). Indeed, convincing *in vivo* examples of genes regulated exclusively by DNA methylation have not emerged and it remains undetermined whether promoter CpG methylation can be the primary and upstream mechanism for directing gene expression patterns.

Despite this assertion, there is limited evidence that the restriction of germline specific genes is mediated primarily by promoter methylation in somatic tissues (Meissner et al., 2008; Rodic et al., 2005; Maatouk et al., 2006). The association between germline specific loci and DNA methylation has been noted in expression studies on hypomethylated ES cells (Fouse et al., 2008) and also in genome wide methylation analyses (Weber et al., 2007). Additionally, Maatouk et al (2006) have shown the temporal activation of three germline specific transcripts in PGCs correlates with promoter demethylation. However, while it has been hypothesised that germline genes could be regulated primarily by DNA methylation, a demonstration of a cause and effect relationship has not emerged. Previous reports have shown largely correlative relationships between expression and promoter methylation status. These studies have not ruled out DNA methylation patterns arising as a consequence of transcriptional activity or as a secondary system mediated by other epigenetic mechanisms. There is thus a problem in attributing causality. Indeed, most studies do not demonstrate that promoter methylation is

required for initiating silencing during development only for maintaining it, which again is compatible with DNA methylation being a secondary maintenance system. A single exception to this is the report by Rodic et al (2005), which convincingly demonstrates *Ant4* requires *de novo* methylation to initiate silencing in differentiating ES cells. However, this report does not rule out other epigenetic systems, including histone modifications, operating upstream of promoter methylation at this locus. Additionally, a study from a second group reports *Ant4* is already highly methylated in undifferentiated ES cells (Suzuki et al., 2007 supp data). Thus, while many commentators often propose germline specific genes as examples of genes regulated primarily by tissue specific methylation, direct evidence for this assertion, rather than correlative evidence, is lacking in the literature. This chapter will aim to comprehensively determine whether promoter methylation functions as the primary and upstream regulatory system for germline-specific gene expression, independently of histone modifications and transcription factor networks. This follows from the previous chapter which noted a striking enrichment of germline-specific genes in a screen for methylation-dependent candidates. Thus here, I will employ a novel cause and effect analysis to investigate the precise role of DNA methylation at the germline-specific candidate loci *Tex19* and *PiwiL2,* identified by my screen (*Table 3.2*).

4.2 The *Tex19* promoter is hypomethylated in expressing cell types but hypermethylated when silenced

Global analyses of the distribution of DNA methylation have shown that CGI promoters (HCP and ICPs) remain largely hypomethylated in all tissues (Weber et al., 2007; Meissner et al., 2008). The *Tex19* promoter is classified as a HCP (516bp Obs/Exp=0.6 GC=59.7%) while the *PiwiL2* promoter is an ICP (424bp Obs/Exp=0.6 GC=61.6%) (*Fig 4.1a & b*). Notably, *PiwiL2* only fails to pass the highly stringent length parameter for HCP classification (500bp), suggesting it is a bona fide CGI promoter. Both of these genes would therefore be predicted to be unmethylated throughout development. However, if promoter CpG methylation were the primary

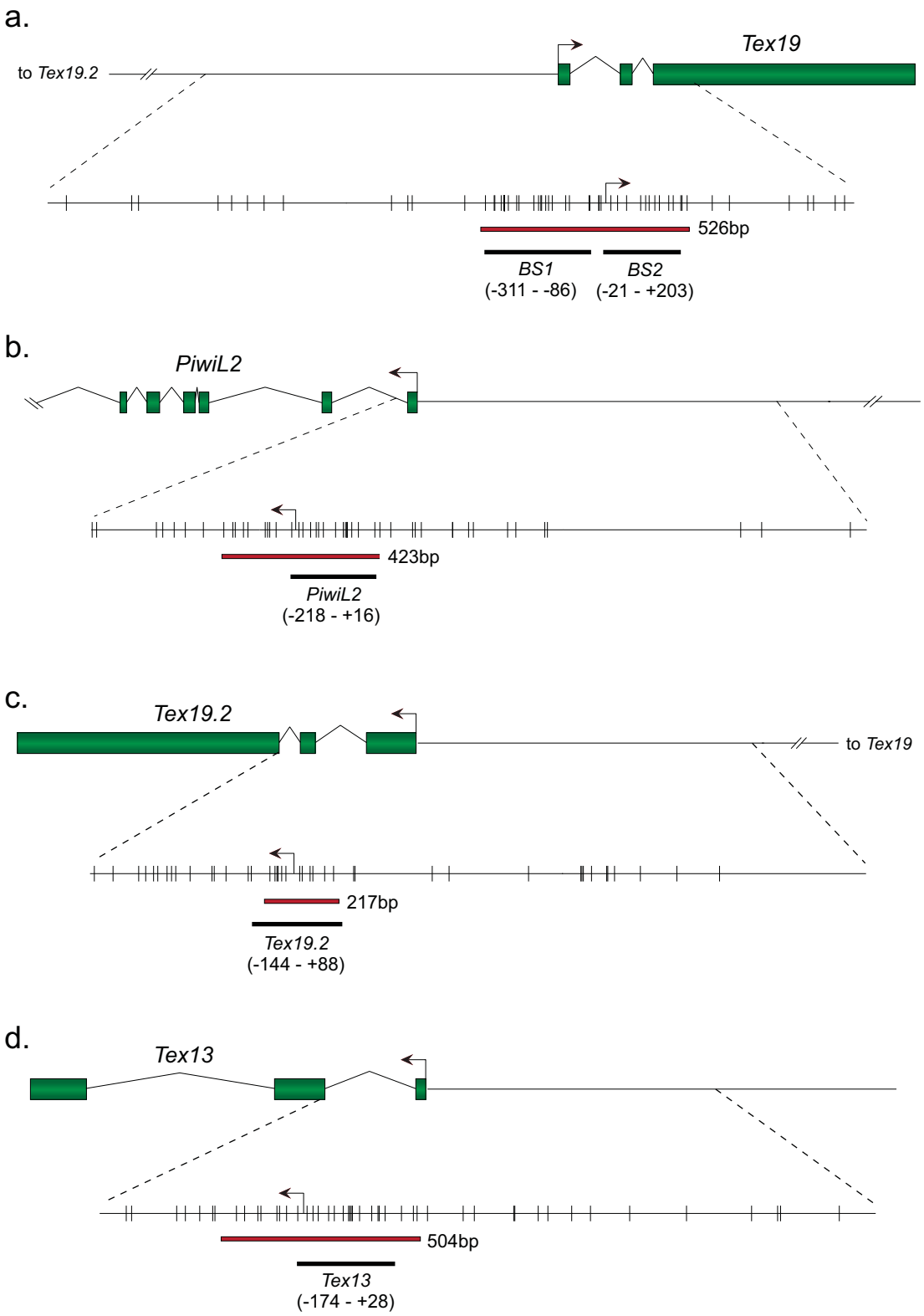


Fig 4.1. **Promoter structures and bisulphite amplified regions of candidate genes.** a.) *Tex19.* b.) *PiwiL2.* c.) *Tex19.2 &* d.) *Tex13.* Green blocks represent exon structure of each gene with CpG content of selected regions magnified below. Each vertical line in the magnified region represents a single CpG dinucleotide with the arrow indicating the transcriptional start site (TSS). Shown in red are CpG islands (CGI) with obs/exp CpG ratio >0.6 and GC content >55%. CGIs should be greater than 500bp to be considered HCPs. CGI size is noted next to each CGI. Shown in black are the regions amplified for bisulphite sequencing analysis with the genomic location relative to the TSS indicated below.

mechanism regulating *Tex19* and *PiwiL2* expression, I would expect to observe heavily methylated promoters in non-expressing tissue-types.

To investigate whether candidate genes are methylated in non-expressing cells, I used the bisulphite sequencing technique to determine the methylation status of CpGs at base pair resolution. Bisulphite treatment distinguishes cytosine from methyl-cytosine (5mC) because cytosine is deaminated to uracil whereas 5mC produces an intermediate that is stable with respect to deamination. Sequencing bisulphite treated DNA thus identifies unmethylated cytosines as thymine residues and, methyl-cytosines as cytosine residues. I initially assayed two regions of the proximal *Tex19* promoter (*Fig. 4.1a*) in pMEFs, P cells and DP cells (performed by *J. Reddington* under my supervision). Here, *Tex19* was found to be heavily methylated (>94%) in non-expressing pMEFs and P cells but as expected, hypomethylated in DP cells (*Fig. 4.2a*). Importantly, this analysis demonstrates that despite being a HCP promoter, which would be predicted to be constitutively hypomethylated, *Tex19* is highly methylated in non-expressing cell-types. This is consistent with promoter CpG methylation potentially being the primary mechanism of silencing *Tex19* expression. Additionally, *Tex19* exhibits markedly reduced promoter methylation in expressing DP cells, supporting the conclusion that these cells are globally demethylated (*Section 3.3*) and consistent with the notion that promoter methylation is critically required to maintain silencing at this loci.

To evaluate whether hypermethylation of the *Tex19* CGI promoter was representative of all the germline specific candidate genes, I examined three more candidate promoters. Similarly to *Tex19,* I found that the *PiwiL2, Tex19.2* and *Tex13* promoters were all heavily methylated in non-expressing pMEFs (*Fig 4.2b*), despite the fact these genes were HCPs or strong ICPs (*Fig 4.1b, c, d & Table 3.2*). I conclude that the candidate germline genes tested here represent a novel sub-class of CGI genes that acquire developmental promoter DNA methylation. It is noteworthy that with the exception of *Tex13* these genes were excluded from published genome-wide methylation analyses and consequently, this is the first data demonstrating that *Tex19, PiwiL2* and *Tex19.2* are novel methylated CGI genes in somatic cells.

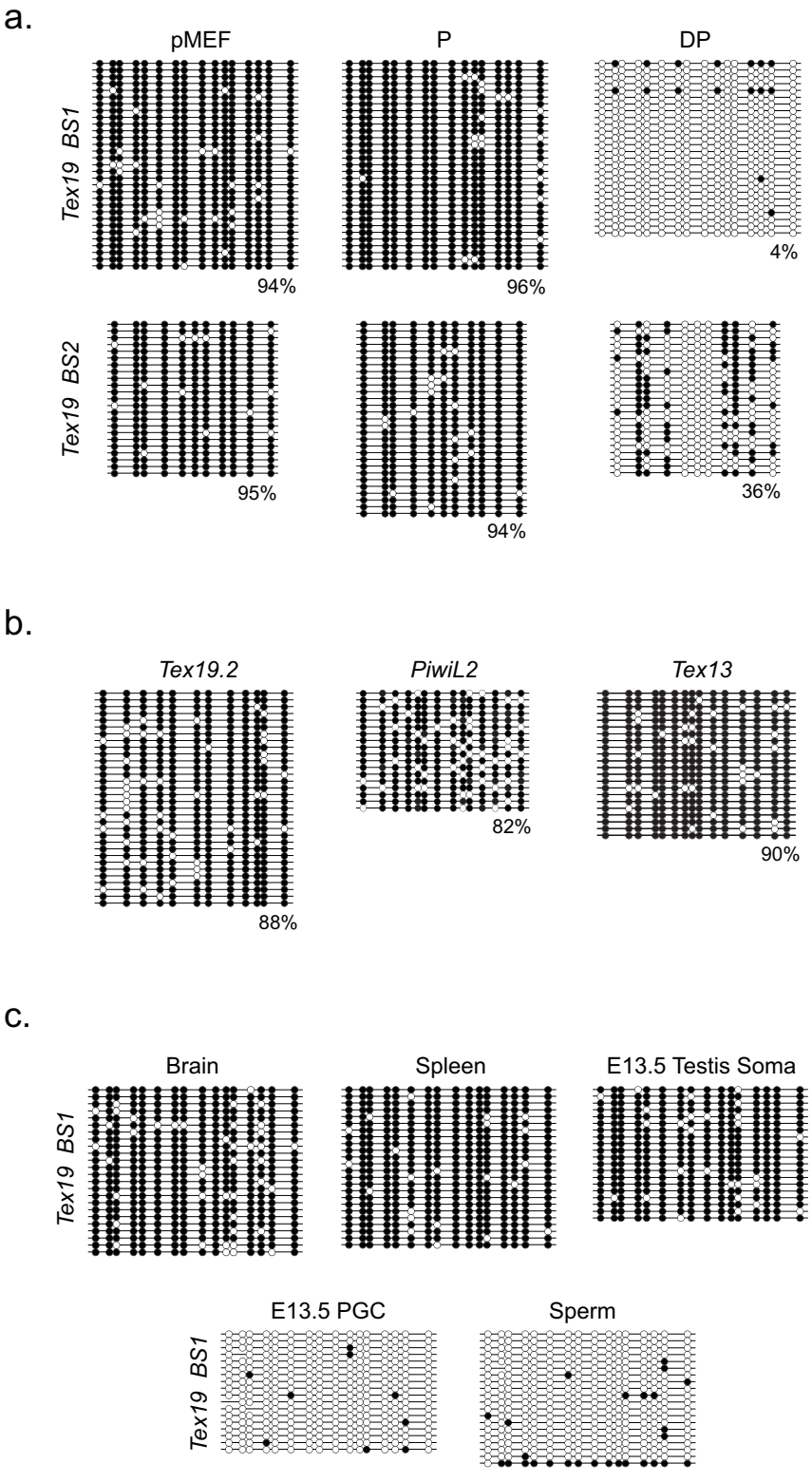


Fig 4.2. **Methylation of candidate gene promoters.** Closed circles represent methylated CpGs, open circles represent unmethylated CpGs. Each horizontal clone represents a unique genomic allele. a.) Bisulphite analysis showing *Tex19* is hypermethylated at both promoter regions (BS1 & BS2) in pMEFs and P cells but hypomethylated in DP cells. b.) Candidate gene CGI promoters are hypermethylated in non-expressing pMEFs, despite being CpG islands. c.) *Tex19* (BS1) is hypermethylated in non-expressing tissues *in vivo* but hypomethylated in expressing PGCs and sperm.

One possibility is that *Tex19* is methylated in all tissues regardless of expression status or that the cell lines used here are not representative of methylation patterns *in vivo*. To investigate this, I examined the *in vivo* methylation of *Tex19* in non-expressing brain, spleen and FACS sorted somatic cells from testis. Additionally, I profiled promoter methylation in expressing purified embryonic day (E)13.5 primordial germ cells (PGC) and mature sperm. Consistent with bisulphite analysis from pMEFs, *Tex19* was highly methylated in non-expressing tissues *in vivo* (>93%), including in somatic cells from the testis. Importantly, *Tex19* was found to be specifically hypomethylated in expressing PGCs and sperm, consistent with *in vivo* demethylation being required for gene activation (*Fig. 4.2c*). These data indicate that promoter CpG methylation strongly correlates with *in vivo* expression patterns and supports the notion that the *Tex19* promoter represents a novel CGI that acquires tissue-specific methylation.

4.3 Tex19 protein is de-repressed in DP cells

RT-PCR analysis of DP cells detected significant de-repression of *Tex19* transcript. To investigate whether *Tex19* transcript in DP cells was full-length and translated I examined Tex19 protein levels. Using a polyclonal α-Tex19 antibody I was able to detect strong staining in DP cells that was blocked by pre-incubation with a Tex19 derived epitope protein, suggesting Tex19 was a unique target (*Fig. 4.3b*). In contrast, P cells exhibited strongly reduced staining (*Fig. 4.3a*). Western blot analysis (performed by *E. Hall* under my supervision) indicated that a band migrating at ~42kDa, the expected size of Tex19, was present specifically in DP cells and testis but not in P cells (*Fig. 4.3c*). I conclude that Tex19 protein is present in DP cells and testis but not in P cells suggesting that demethylation of the *Tex19* promoter is sufficient to promote expression and translation of full length Tex19 transcript.

4.4 *Tex19* and *PiwiL2* promoters drive expression of reporter genes in somatic cells

The data presented here suggests that *Tex19* and *PiwiL2* are methylated in cultured non-expressing cells (*Section 4.2*), that experimental demethylation induces

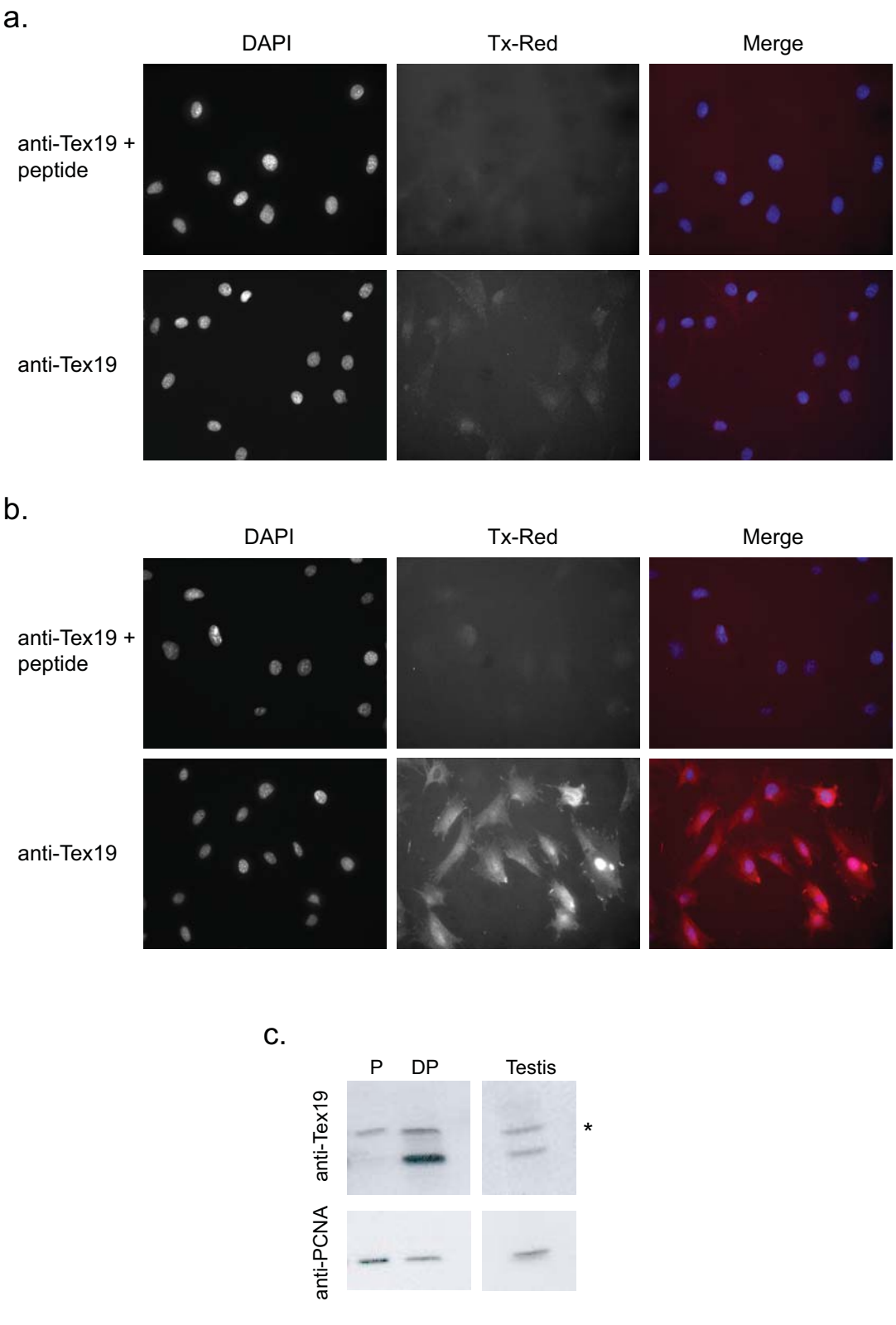


Fig 4.3. **Tex19 protein is expressed in DP cells but not P cells.** a.) Tex19 staining cannot be significantly detected in P cells by immunofluoresence. DAPI shown in blue, stains nuclei. b.) Strong Tex19 staining (red) is evident in DP cells (bottom panel) which is abrogated by pre-incubation with a Tex19-specific peptide. c.) Western blot demonstrating that Tex19 can be specifically detected in DP cells and testis but not in P cells. Asterik denotes non-specific band. PCNA was used as a loading control

expression (*Section 3.5 & 3.6*) and that the *in vivo* promoter methylation status correlates with protein expression (*Section 4.3*). As part of a strategy to attribute functional significance to DNA methylation at candidate loci, I cloned the proximal promoters of *Tex19, PiwiL2* and *Tex13* into luciferase reporter constructs (pGL3-basic). Additionally I cloned the promoters of three control germline-specific genes (*Rnh2, Pramel1* and *Tex11*), that each failed at least one criterion required to be considered as methylation-dependent candidates (*Section 3.4*). For all constructs I cloned the promoter region from approximately 1kb upstream to up to 250bp downstream of the transcriptional start site (TSS) of the primary isoform (RefSeq). Transient transfection of these constructs into mouse Neuro2a cells or human 293T cells demonstrated that all three bona fide candidate gene promoters (*Tex19, PiwiL2, Tex13)* could drive strong expression (up to 130-fold) of the luciferase reporter compared to vector only (*Figure 4.4*). Indeed *Tex19* was observed to drive comparable reporter expression to the highly active CMV promoter (~25% of maximal CMV) (data not shown). This indicates that unmethylated *Tex19* is an inherently strong promoter. In contrast, two of the three control germline promoters (*Rnh2* and *Tex11*) and the negative control, *Oct3/4,* were unable to promote significant reporter expression above background (*Fig 4.4*).

This analysis reveals two important points. Firstly, all three candidate genes tested here, *Tex19, PiwiL2* and *Tex13,* were able to drive strong reporter expression in terminally differentiated mouse and human somatic cells. This demonstrates that these genes do not have a critical reliance on germline-specific transcription factors (TF) for expression. Instead, it is likely that ubiquitous and general TFs are driving the observed expression from these promoters. Secondly, despite the fact that the *Tex19, PiwiL2* and *Tex13* promoters are able to drive strong transcription in Neuro2a and 293T cells, the endogenous loci are silenced in both cell-types. This indicates that the endogenous loci critically rely on an epigenetic mechanism to maintain their silenced state. Both of these conclusions are consistent with promoter CpG methylation being the primary and exclusive mechanism regulating expression of these candidate genes.

In contrast, the germline control genes that were predicted not to be regulated primarily by DNA methylation exhibited only weak or background expression in these experiments (*Fig 4.4*). It would be predicted these genes partly rely on lineage specific TFs for their expression, as has been demonstrated for *Oct3/4 (*Yoem et al., 1996; Gu et al., 2005). At these loci, DNA methylation may contribute to the regulatory mechanism and/or demethylation may be required to 'poise' these genes for activation in the presence of appropriate TFs but CpG methylation cannot be the only or primary mechanism of regulating their transcription. The exception to this is *Pramel1,* which is expressed strongly in these reporter assays. Interestingly, *Pramel1* was considered a germline-specific control gene here because it was marginally below the threshold in 3T3-recovery cells, although it was activated in DP cells and P-aza cells. In contrast, the other control genes such as *Oct3/4* and *Rnh2* were not activated by 5-aza dC treatment. *Pramel1* may therefore represent a gene that was excluded from the candidate list due to the highly stringent thresholds employed here and may indeed critically rely on epigenetic mechanisms for regulation. In summary, this reporter analysis demonstrates that *Tex19* and *PiwiL2* do not require germline specific TFs and suggests that an epigenetic mechanism exclusively mediates the expression pattern of the endogenous loci irrespective of cellular lineage.

To investigate the role of distinct *cis* elements in the *Tex19* and *PiwiL2* promoters I generated a set of deletion constructs from each promoter. Following transient transfection into 293T cells, I observed no significant differences between the two *PiwiL2* promoter regions (*Fig 4.5*). However, transfection of the *Tex19* constructs identified discrete functional regions within the promoter. Interestingly, the proximal downstream +58 - +198 region strongly repressed transcription by ~7-fold compared to the empty vector background level. In contrast, the proximal upstream –224 - +58 construct was able to drive strong expression ~245-fold above background and ~2.5-fold higher than the entire promoter construct. Removal of the region upstream of -304bp had no significant effect on transcription (compare -1034 - +224 and –304 - +198) indicating that all the functional *Tex19 cis* motifs reside proximal to the TSS or outside of the regions assayed here. These data indicate that *Tex19* contains at least two functional elements, which strongly activate (–224 - +58) or weakly

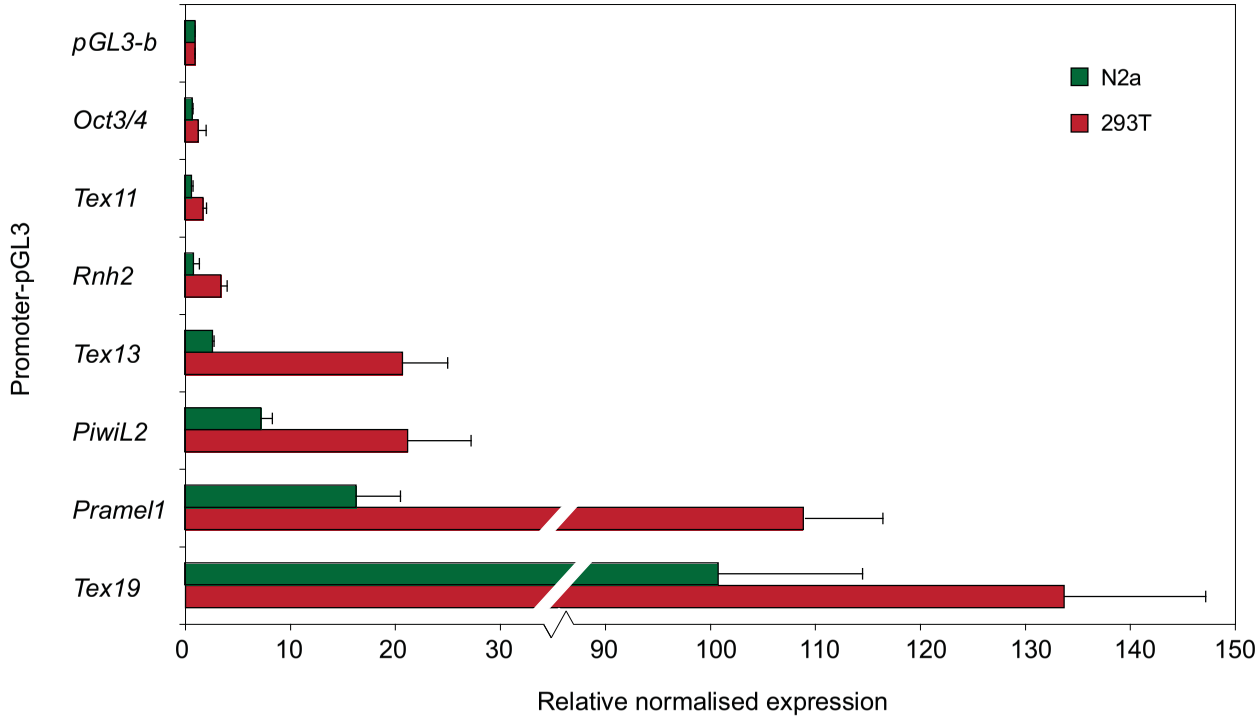


Fig. 4.4. **Reporter analysis of germline-specific gene promoters.** Proximal gene promoters of candidate genes (*Tex19, PiwiL2, Tex13*), control germline genes (*Rnh2, Pramel1, Tex11*) and negative control (*Oct3/4*) were cloned into pGL3-basic and transiently transfected into Neuro2a or 293T cells. Shown is firefly luciferase reporter activity normalised to a co-transfected renilla luciferase. Expression is relative to empty vector (pGL3-b) which is set at 1.

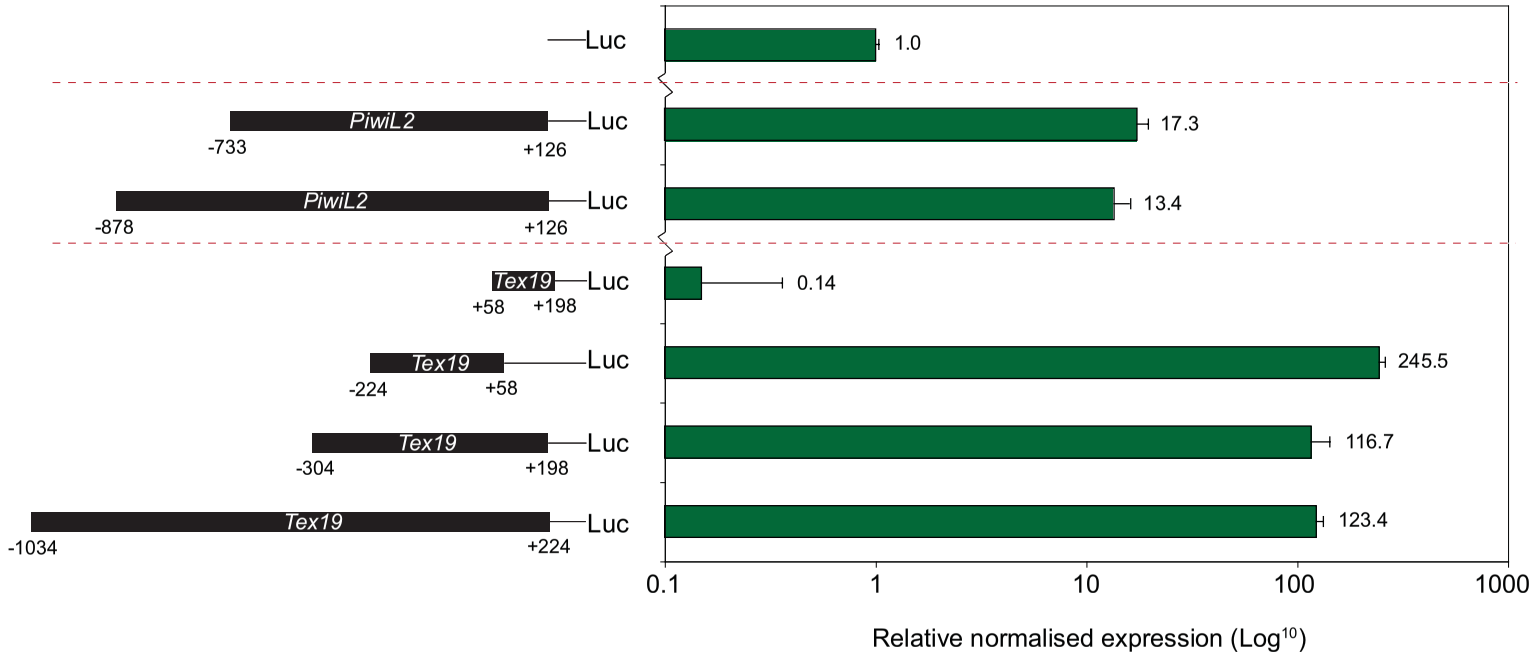


Fig 4.5. **Deletion constructs of the *Tex19* and *PiwiL2* promoters.** Each promoter construct was transiently transfected into 293T cells and reporter expression was normalised to co-transfected renilla luciferase. Shown is expression relative to empty vector (pGL3-b), which is set at 1. Scale is calibrated to log10. Promoter regions are relative to transcriptional start site (TSS) (Refseq).
:

repress (+58 - +198) transcription, respectively (*Fig 4.*5). When both elements are present (–304 - +198) strong expression is still observed but reduced by approximately 2.5-fold suggesting that the *cis*-activating motif is dominant but influenced by the *cis*-repressor domain. Analysis of each promoter region with the Transcription Elements Search System (TESS) programme failed to identify any significant (*Poisson* model p = <0.05) differences in transcription factor sites between the *Tex19* activating (–224 - +58) and repressing (+58 - +198) regions. Electrophoretic mobility shift assays (EMSA) will be required here in future to identify protein factors that bind each distinct *Tex19 cis* element and direct their distinct properties. I conclude that the strong *Tex19* expression in somatic cells is predominantly driven by the –224 - +58 region of the promoter but partially constrained by an adjacent repressor +58 - +198 motif.

4.5 Promoter CpG methylation silences the *Tex19* and *PiwiL2* reporters

Reporter analysis of unmethylated *Tex19* and *PiwiL2* promoters suggested that transcription factors are not limiting for expression of the endogenous loci and that an epigenetic mechanism must therefore exclusively silence their transcription. To investigate whether DNA methylation is capable of repressing *Tex19* and *PiwiL2* activity, I *in vitro* methylated the reporter constructs using *Sss.I* methylase, which methylates all CpG sites and transiently transfected methylated or mock methylated constructs into 293T cells. Here, expression of *Tex19* was repressed ~105-fold and *PiwiL2* was repressed ~145-fold compared to mock methylated controls. In contrast, methylation of the empty reporter construct only repressed transcription by 17-fold (*Fig 4.6a*). This translates into a net repression of 6-fold (*Tex19*) and 8.5-fold (*PiwiL2*), which suggests CpG methylation can repress expression from these promoters significantly below background effects.

A drawback of the above experimental method is that the entire construct becomes CpG methylated including the reporter coding region. This could potentially have indirect effects independent of promoter methylation. To address this, I used the

experimental scheme shown in *Fig 4.6b* to generate reporter constructs with only the promoter region methylated. Using this approach I investigated the effect of promoter-specific methylation on each *Tex19* deletion construct. Here I found that methylation of the +58 - +198 repressor region had no significant effect on transcription compared to the mock methylated construct (*Fig 4.6c*). This is consistent with this domain already functioning as a transcriptional repressor element independent of methylation. In contrast, methylation of the –224 - +58 activator region strongly repressed transcription by 12-fold compared to mock methylated controls (*Fig 4.6c*). This assay clearly demonstrates that promoter CpG methylation is sufficient to promote strong repression of *Tex19* and is consistent with DNA methylation mediating silencing of endogenous *Tex19.* It is noteworthy that these assays are conducted with reporters that are outwith a chromatin environment. Therefore the degree of silencing attainable by promoter DNA methylation will be limited by context. These assays merely provide proof of principle that CpG methylation can have a significant effect on *Tex19* repression and do not provide a quantitative measure of the effect of endogenous promoter methylation, which may be considerably more effective. Thus, these data suggest that in a chromatin context, methylation of the CpG dense *Tex19* promoter would induce significant transcriptional silencing.

4.6 *Tex19* and *PiwiL2* are not de-repressed in Lsh-/- MEFs

To further investigate the direct link between *Tex19* expression and promoter hypomethylation I used globally hypomethylated *Lsh*-null MEFs. Lsh is a putative SWI/SNF chromatin remodelling protein that is essential for maintaining DNA methylation at specific transposable elements, *Hox* genes and imprinted loci, probably via its interaction with the *de novo* methyltransferases (Zhu et al., 2006; Xi et al., 2007; Fan et al., 2005). I predicted that the absence of Lsh would lead to hypomethylation and activation of *Tex19.* Importantly, this would exclude the possibility that de-repression of *Tex19* in DP and P-aza cells occurred due to loss of Dnmt1 protein *per se*, which has been reported to repress transcription independently of its methyltransferase activity (Fuks et al., 2000; Robertson et al., 2000; Dunican et

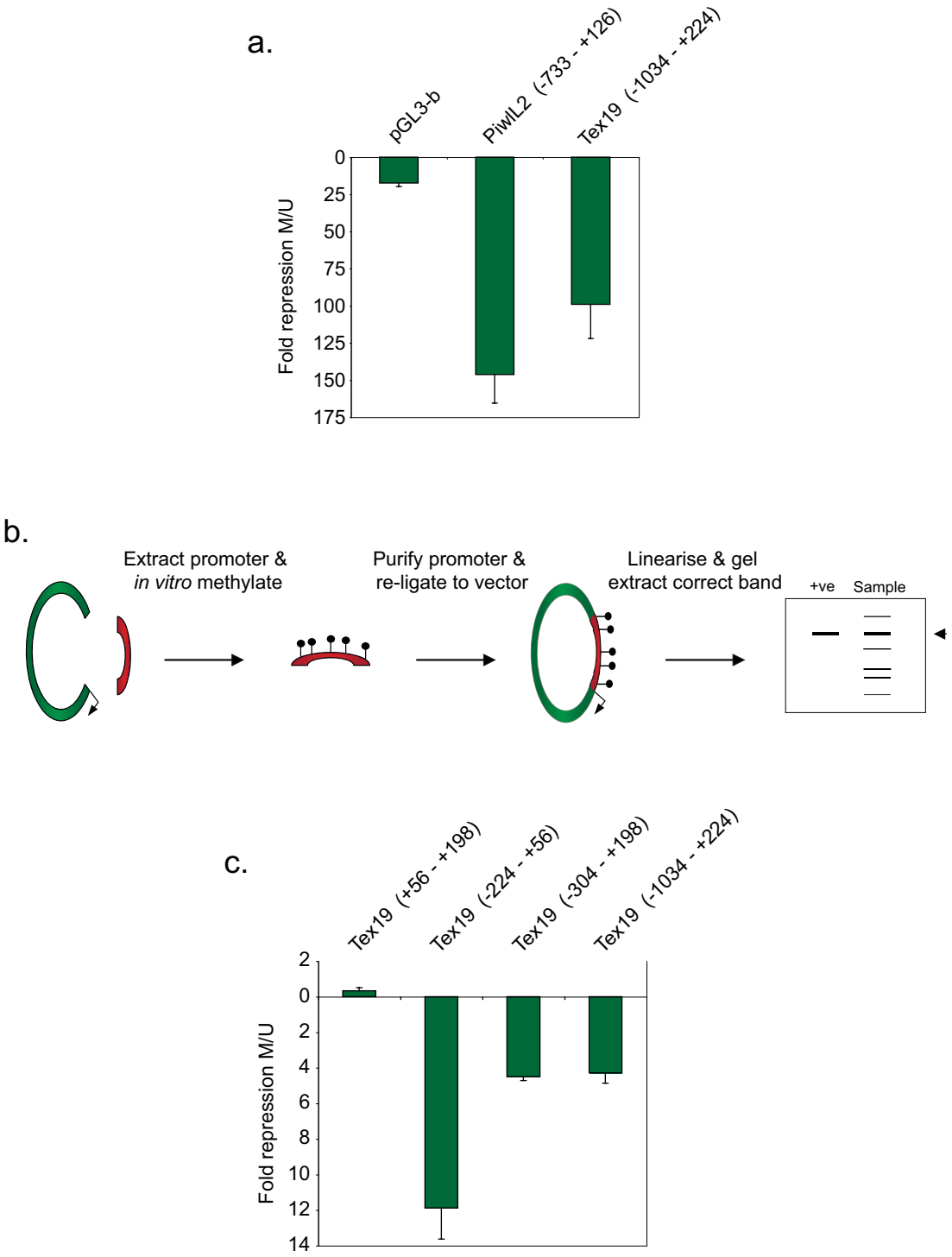


Fig 4.6. **Promoter methylation of the *Tex19* and *PiwiL2* reporters represses transcription.** a.) *PiwiL2* and *Tex19* reporters were *in vitro* methylated or mock methylated and transiently transfected into 293T cells. Shown is relative repression as determined by the ratio of methylated reporter expression (M) to unmethylated reporter expression (U). b.) Schematic of method used to generate reporter constructs with promoter specific methylation. Briefly, the promoter region was digested out with restriction endonucleases, *in vitro* methylated and re-ligated to the unmethylated reporter backbone. This was linearised, size seperated to extract the limited number of single promoter region insertions and purified. c.) Relative repression of *Tex19* deletion constructs specifically methylated only in the promoter region relative to mock methylated controls

al., 2008). In the present study, digestion of genomic DNA from Lsh-deficient cells with the methyl-sensitive *HpaII* restriction enzyme confirmed these cells were globally hypomethylated (data not shown). Surprisingly, bisulphite analysis of *Tex19* and *Tex19.2* demonstrated that these loci retain fully methylated promoters in *Lsh*-null MEFs (*Fig 4.7a*). In contrast, *Oct3/4* was highly demethylated, confirming that *Lsh*-null cells were hypomethylated and that promoter regions were susceptible to demethylation (*Fig 4.7b*). RT-PCR analysis of several candidate germline genes revealed no de-repression of transcripts in *Lsh*-null cells, consistent with promoter specific demethylation being required for activation of these genes (*Fig 4.7*c). In contrast despite promoter hypomethylation, *Oct3/4* was not activated in Lsh-deficient cells, probably because this locus is regulated by multiple levels of epigenetic and trans-acting systems in addition to DNA methylation (*Fig 4.7d*) (Feldman et al., 2006). Notably, the critical reliance on Lsh to maintain promoter methylation at *Oct3/4* was a novel observation at the time of these experiments (2008) but has recently been reported elsewhere (Xi et al., 2009).

While *Tex19* unexpectedly retained methylation in globally hypomethylated *Lsh*-null cells, this analysis revealed several interesting observations. Firstly, despite global demethylation in Lsh deficient cells, including at *Oct3/4* and three other non-candidate promoters tested (data not shown), *Tex19* maintains normal promoter methylation. This suggests that, unlike most loci examined here and elsewhere (Zhu et al., 2006; Fan et al., 2005), *Tex19* does not rely on Lsh to retain CpG methylation and instead utilises a distinct and/or combination of mechanisms to target and maintain promoter methylation. This property could potentially reflect the critical reliance on promoter CpG methylation to maintain *Tex19* silencing. Secondly, methylated *Tex19* is not activated in a globally demethylated environment. This indicates that *Tex19* is not activated by the secondary or indirect effects of cellular demethylation and supports the argument that this locus is directly regulated by promoter CpG methylation *in cis*. However, these experiments cannot rule out loss of Dnmt1 protein *per se* activating *Tex19* in DP and P-aza cells.

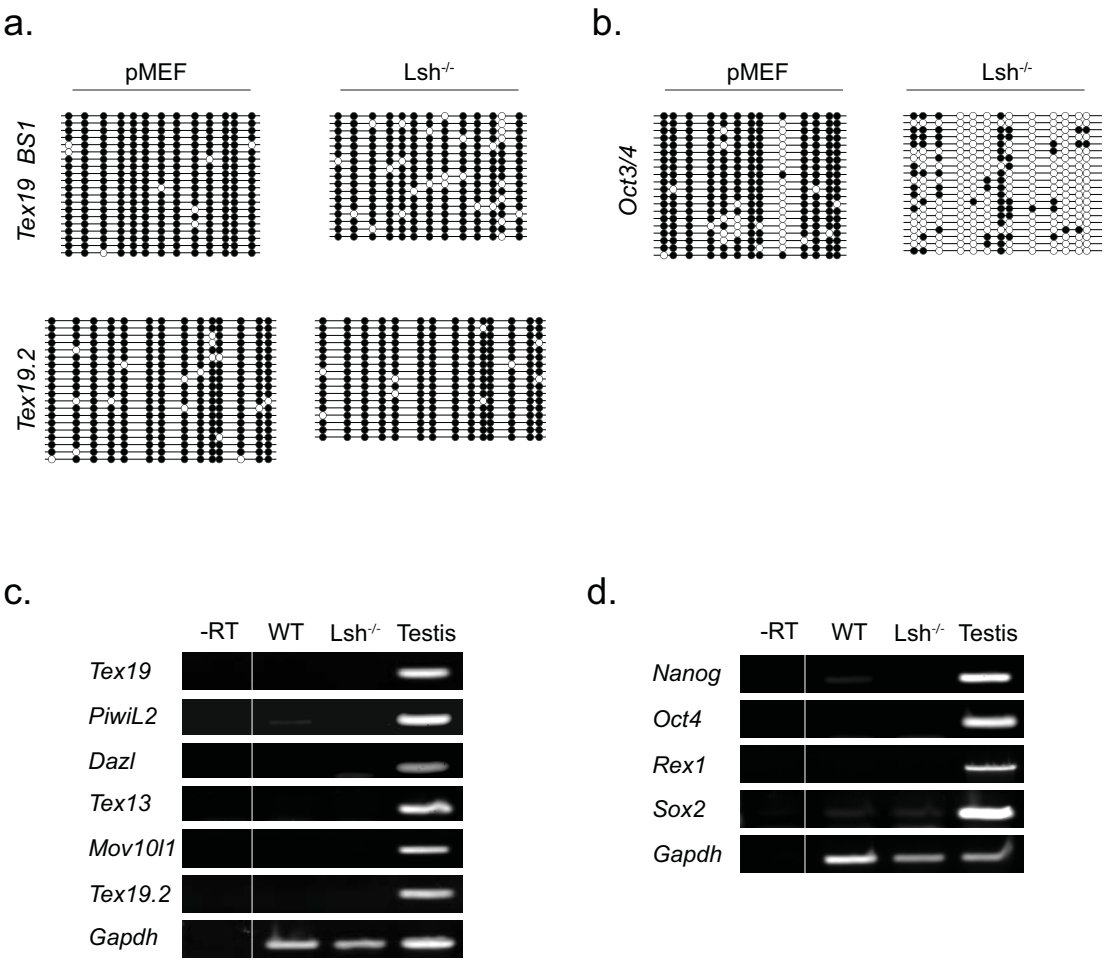


Fig 4.7. **Germ-line candidate genes are not demethylated and activated in hypomethylated *Lsh*-deficient MEFs.** a.) Bisulphite methylation analysis of candidate genes *Tex19* and *Tex19.2* in primary MEFs (pMEF) and *Lsh*-null MEFs. b.) Methylation of control *Oct3/4* in pMEFs and *Lsh*-null MEFs. c.) RT-PCR analysis of candidate gene expression and of d.) pluripotency-associated gene expression.

4.7 *Tex19* & *PiwiL2* acquire promoter methylation and silencing during ES cell differentiation

To investigate the role of promoter CpG methylation at *Tex19* and *PiwiL2* during cellular differentiation, I induced ES cells to differentiate into embryoid bodies (EB). The central concept of EB culture is to mimic the embryonic process of germ layer formation from the ICM. Here, aggregated ES cells develop into a cavity structure that includes multiple lineages derived from the primitive ectoderm and visceral endoderm. This structure, while less organised than an embryo, has been shown to contain antero-posterior polarity and form a primitive streak-like region (ten Berge et al., 2008). EB differentiation thus closely parallels *in vivo* development (Nishikawa et al., 2007). In the present study, RT-PCR analysis demonstrated that *Tex19* and *PiwiL2* were strongly expressed in undifferentiated mouse E14 ES cells, consistent with the GNF microarray database, which indicates these genes are expressed in zygotes and blastocysts (GNF Symatlas - http://biogps.gnf.org) and previous reports (Kuntz et al., 2008; Silva et al., 2009). However, both genes underwent progressive temporal repression upon induction of EB differentiation, with silencing observed by day 4 (*Fig 4.8a*).

If promoter methylation were the primary mechanism directing silencing of these loci, it would be predicted that DNA methylation would accumulate prior to or coincident with gene silencing. To test this, I evaluated *Tex19* and *PiwiL2* promoter methylation dynamics during EB differentiation. Here, *Tex19* was hypomethylated in undifferentiated ES cells (day 0) but acquired significant methylation by day2 (57%) and was fully methylated (87%) by day 4 of EB formation (*Fig 4.8b*). *PiwiL2* followed similar developmental CpG methylation dynamics (*Fig 4.8c*). The small number of methylated clones at day 0 probably represents a sub-population of partially differentiated ES cells. These data strongly support the conclusion that promoter methylation is targeted to *Tex19* and *PiwiL2* coincident (at least) with gene silencing (*compare Fig 4.8a with Fig 4.8b & c*). In contrast, studies on the developmental silencing of *Oct3/4* have shown that DNA methylation accumulates at the promoter after transcriptional silencing has been imposed and is therefore a

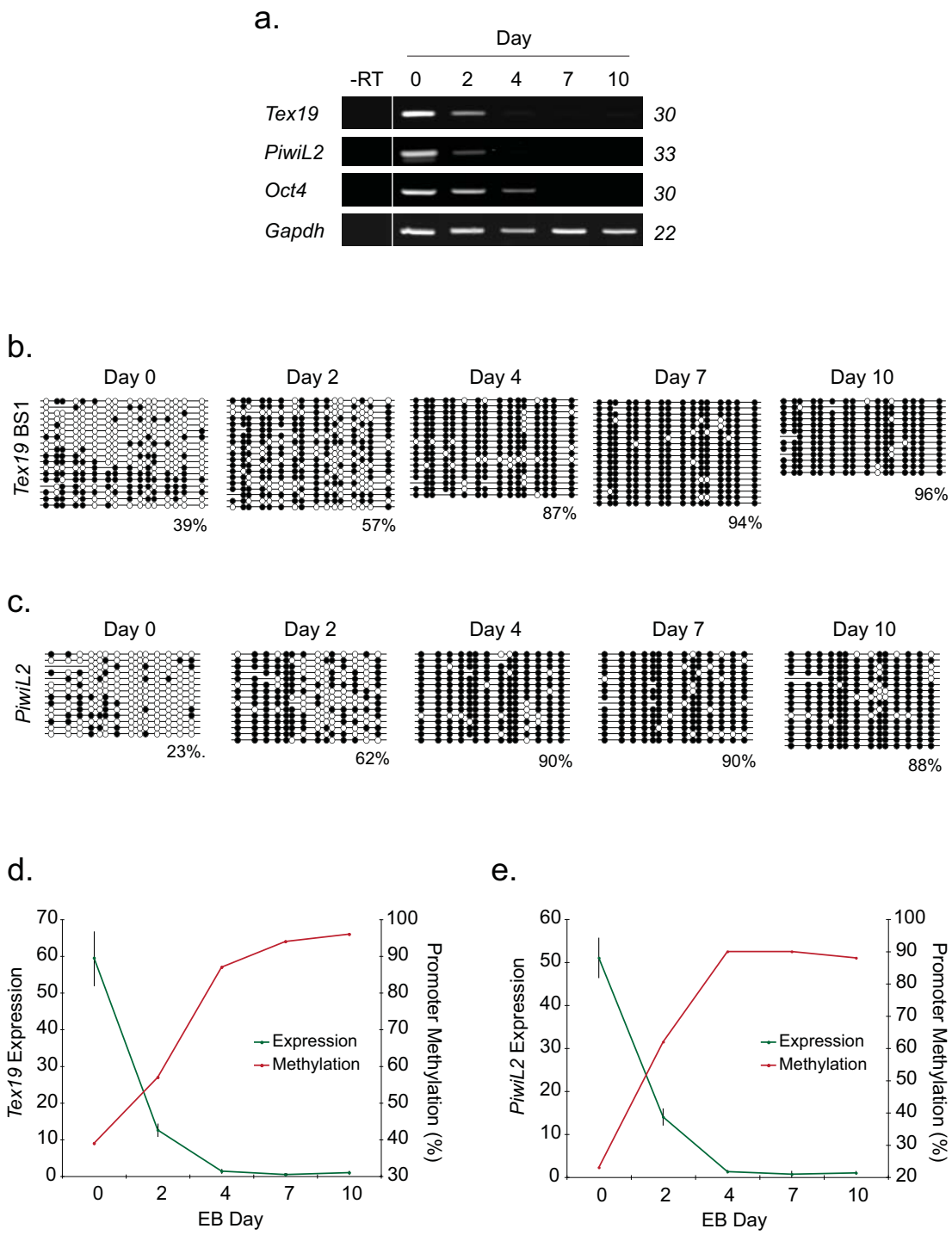


Fig 4.8. **Promoter methylation dynamics during EB differentiation.** a.) RT-PCR analysis of *Tex19* and *PiwiL2* expression during ES cell EB differentiation. *Oct3/4* demonstrates EBs are differentiating. Right panel shows cycle number. b.) *Tex19* and c.) *PiwiL2* promoter methylation dynamics during EB formation. Percentage methylation shown on bottom right. Methylation accrues at gene promoters before or coincident with transcriptional silencing. d.) & e.) Comparison of qRT-PCR expression and promoter methylation during EB differentiation shows an inverse correllation at d.) *Tex19* & e.) *PiwiL2.* Expression is relative to day 10, which is set to 1 and normalised to *Gapdh*. Error bars express S.E.M.

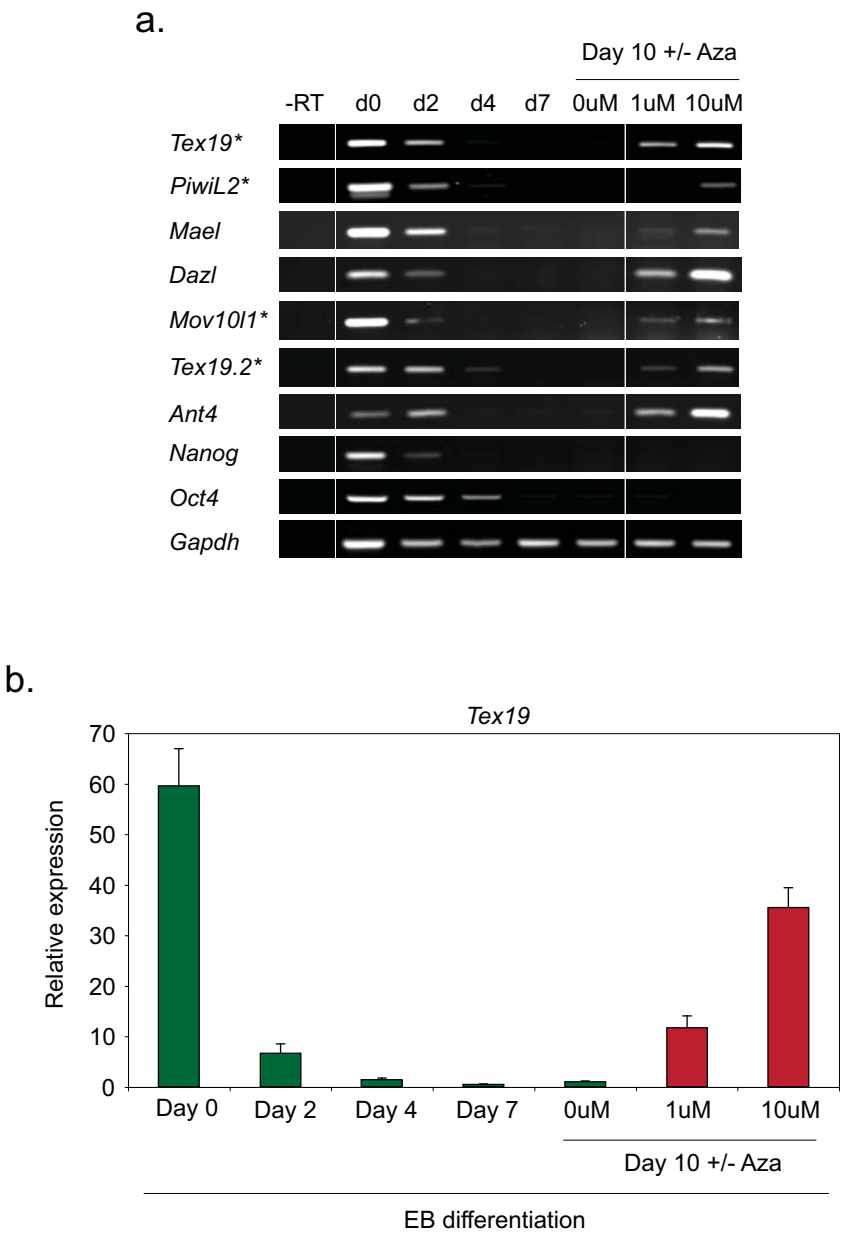


Fig 4.9. **5-aza dC re-activates silenced candidate genes after EB differentiation.** a.) RT-PCR expression analysis of candidate (asterik) and germline genes demonstrates they are silenced during EB formation and subsequently re-activated by 5-aza dC (Aza) treatment for the final 72hrs of EB differentiation. Control pluripotentcy genes *Oct3/4* and *Nanog* are not re-activated. b.) qRT-PCR of *Tex19* expression in EBs +/- 5-aza dC. Expression is relative to day 10 control -Aza, which is set to 1 and normalised is *Gapdh*

secondary repressive system at this locus (Feldman et al., 2006). Here, the data is consistent with promoter methylation being the primary mechanism of *Tex19* and *PiwiL2* silencing as methylation occurs in parallel with repression. To quantitatively confirm the dynamic relationship between gene expression and CpG methylation I used qRT-PCR. This analysis clearly demonstrates a direct inverse relationship between acquisition of promoter methylation and silencing of *Tex19* and *PiwiL2* (*Fig 4.8d & e*)

To further analyse promoter methylation dynamics, I used 5-aza dC to study the effects of demethylation on candidates genes silenced during EB differentiation. Here, I treated EBs with high (10uM) or low (1uM) 5-aza dC for the final 3 days of EB culture. Notably, all the candidate genes tested, including *Tex19* and *PiwiL2,* were silenced during differentiation but de-repressed by addition of 5-aza dC. In contrast, mock treated EBs (0uM) maintained repression of candidate genes (*Fig 4.9a & b*). This analysis suggests that candidates are not being silenced through changes in transcription factor availability during differentiation, as they can be readily reactivated by demethylation. Additionally 5-aza dC treatment was insufficient to de-repress the pluripotency factors *Nanog* and *Oct3/4,* consistent with these loci acquiring several tiers of repressive modifications, in contrast to *Tex19* and *PiwiL2,* which critically rely on CpG methylation for silencing *(Fig 4.9a*). In summary, this analysis has demonstrated that *Tex19* and *PiwiL2* are methylated before or in parallel with transcriptional silencing during EB differentiation and that this silencing can be relieved by promoter demethylation.

4.8 *Tex19* and *PiwiL2* rely on *de novo* methylation mediated by Dnmt3b for silencing during ES cell differentiation

To conclusively demonstrate that DNA methylation is both causal and necessary to direct gene silencing at *Tex19* and *PiwiL2* I used ES cells deficient for Dnmt3a and Dnmt3b. These enzymes have been proposed to be the primary mediators of *de novo* methylation during murine embryonic development (Okano et al., 1999). Consequently, ES cells lacking both Dnmt3a and Dnmt3b have undetectable *de novo*

methylation activity. Despite this, early passage *Dnmt[3a 3b]*-null ES cells retain global methylation, due to Dnmt1 activity, and can differentiate *in vitro* (Jackson et al., 2004; Gilbert et al., 2007). Here, I differentiated early passage *Dnmt[3a 3b]*-null ES cells into embryoid bodies in parallel with parental (wild-type) J1 ES cells. As observed with E14 ES cells, *Tex19* and *PiwiL2* were strongly silenced during EB differentiation of wild-type J1 ES cells (*Fig 4.10a*). However intriguingly, *Tex19* and *PiwiL2* were unable to impose transcriptional silencing during differentiation in the absence of Dnmt3a and Dnmt3b (*Fig 4.10b*). In contrast, *Oct3/4* was silenced comparatively to wild-type cells, indicating *Dnmt[3a 3b]*-null ES cells proceed with normal differentiation, as has been previously reported (*Fig 4.10b*) (Gilbert et al., 2007; Li et al., 2007). To confirm that *Tex19* and *PiwiL2* were intrinsically unable to silence expression and not just delayed in *Dnmt[3a 3b]*-null ES cells, I cultured EBs for an extended period. Here strong expression of *Tex19* and weak expression of *PiwiL2* could still be detected, even after 15 days differentiation, indicating that both genes are unable to impose full gene silencing without *de novo* methylation (data not shown). The failure of *Tex19* and *PiwiL2* to silence expression strongly supports the notion that *de novo* promoter methylation is the critical and primary mark mediating repression at these loci. Additionally, this evidence supports the conclusion that Dnmt1 protein *per se* is not the causal factor repressing *Tex19* and *PiwiL2* in DP and P-aza cells, as Dnmt1 is present in expressing day 10 *Dnmt[3a 3b]*-null EBs.

Notably, qRT-PCR analysis suggested *PiwiL2* transcript levels were repressed ~7-fold by day 10 in *Dnmt[3a 3b]*-null EBs. While this is considerably less than wild-type embryoid bodies (51-fold), it still represents significant gene repression. This may reflect a change in the availability of transcription factors or may indicate that other epigenetic systems have a secondary influence on *PiwiL2* regulation. Interestingly *Ant4,* which is the only other gene to my knowledge reported to rely on *de novo* methylation for silencing (*see Section 4.1* & Rodic et al., 2006), exhibited similar moderate repression to *PiwiL2* in *Dnmt[3a 3b]*-null EBs. Thus while promoter methylation is critically required for silencing of *PiwiL2* and *Ant4*, my data suggests that the precise level of expression of these genes may also be influenced by other factors. In contrast, *Tex19* expression remained completely unaffected during

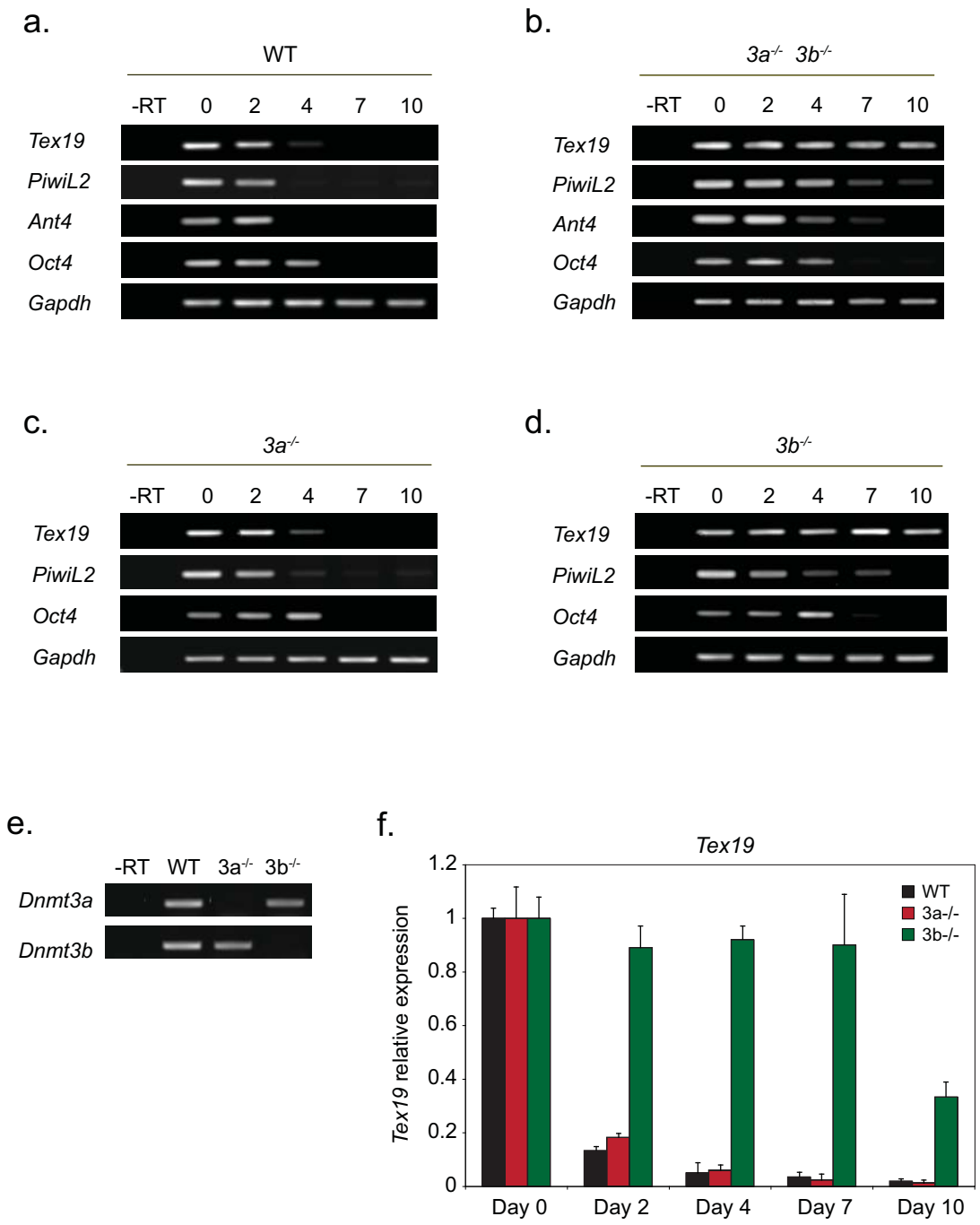


***Fig 4.10. Tex19* and *PiwiL2* are not silenced in differentiating Dnmt3-null ES cells.** RT-PCR analysis of *Tex19* and *PiwiL2 expression* during EB differentiation in a.) WT (J1) and b.) early passage *Dnmt [3a 3b]*-null ES cells. *Tex19* and *PiwiL2* are silenced during WT ES cell differentiation but *Dnmt [3a 3b]*-null ES cells are unable to impose silencing on *Tex19* and *PiwiL2* is severaly delayed. Note *Oct3/4* silencing is unaffected by the absence of *Dnmt [3a 3b].* c.) & d.) RT-PCR analysis of *Tex19* and *PiwiL2* expression in single *Dnmt3a*-null or *Dnmt3b*-null EBs. e.) RT-PCR genotyping of *Dnmt3a* and *Dnmt3b* single null ES cells. f.) qRT-PCR analysis of *Tex19* expression during EB differentiation in wild-type, *Dnmt3a*-null and *Dnmt3b*-null ES cells relative to parental undifferentiated ES cells (day 0), which are set to 1. Expression is normalised to *Gapdh*. Error bars represent the propergated S.E.M.

differentiation of *Dnmt[3a 3b]*-null EBs, strongly suggesting that promoter methylation is the crucial and exclusive mechanism of regulating this gene.

To test whether either of the Dnmt3 family members played a predominant role in targeting methylation to *Tex19* or *PiwiL2* I used ES cells singly lacking Dnmt3a or Dnmt3b (*Fig 4.10e*). RT-PCR analysis of EB differentiated *Dnmt3a*-null ES cells, showed that *Tex19* and *PiwiL2* imposed silencing comparably to wild-type parental J1 ES cells (*Fig 4.10c*). However, differentiation of *Dnmt3b*-null ES cells clearly demonstrated that *Tex19* was unable to silence transcription and that *PiwiL2* repression was strongly delayed (*Fig 4.10d*). Importantly, both cell lines were capable of differentiation, as *Oct3/4* was repressed similarly to wild-type ES cells. qRT-PCR confirmed that wild-type and *Dnmt3a*-null ES cells imposed strong repression of *Tex19* but *Dnmt3b*-null ES cells were unable to impose such silencing (*Fig 4.10f*). This data suggests that Dnmt3b is the critical enzyme responsible for directing promoter methylation to *Tex19* and thus mediating silencing at this locus. Furthermore, the incomplete silencing of *PiwiL2* in Dnmt3b-null cells suggests Dnmt3b is a crucial component required for repression of *PiwiL2*. In summary these data provide strong cause and effect evidence that promoter methylation, targeted by Dnmt3b, is the primary and exclusive mechanism of regulating *Tex19* expression. In contrast, this data is consistent with *de novo* DNA methylation functioning as a crucial component of the regulatory system of *PiwiL2*.

4.9 Silenced Tex19 is not marked by histone modifications

While the data presented here strongly supports a primary and causal role for DNA methylation in regulating *Tex19*, one possibility is that histone modifications contribute to a secondary or targeting regulatory system at the locus. To evaluate the role of histone modifications at *Tex19,* I initially interrogated the genome-wide histone modification database generated by Mikkelsen et al (2007). This indicated that the promoter region of *Tex19* (~1.5kb around TSS) was not enriched for any of the histone modifications examined (*Appendix 4*). Indeed, the *Tex19* promoter was highly *depleted* of histone modifications. Importantly, there was no difference in the

modification profile between expressing ES cells and non-expressing pMEFs, strongly suggesting that histone modifications do not play an important role in promoting the differential expression observed between these cell types.

To experimentally validate the relative depletion of histone modifications at *Tex19*, I performed native chromatin-immunoprecipitation (n-ChIP) from P and DP cells at the two regions of *Tex19* shown in *Fig 4.11a*. Of particular interest are the modifications present in non-expressing P cells which could potentially contribute to silencing of *Tex19*. The absence of histone marks in P cells would render promoter CpG methylation as the likely exclusive mechanism directing gene silencing. ChIP analysis of the repressive polycomb mark H3K27me3 demonstrated that this modification is absent from, or only weakly deposited, at *Tex19* in P cells as determined by the *B-actin* negative control and *Oct3/4* positive control (*Fig 4.12a*). Furthermore, ChIP of the activating mark H3K4me2, which has been reported to structurally inhibit *de novo* methylation (Ooi et al., 2007), indicated that this modification is also depleted from *Tex19* in P cells compared to the negative *Oct3/4* control but increased moderately in DP cells (*Fig 4.12b*). Importantly, these analyses suggest that neither H3K27me3 nor H3K4me2 is significantly present at *Tex19* in non-expressing P cells and therefore cannot be involved in regulating *Tex19* silencing. Further ChIP analysis of the H3K9ac mark and the histone variant H2A.Z suggested that neither of these epigenetic states were significantly enriched in P cells either (*Fig 4.12a & b*). This is consistent with the Mikkelsen global modification database (*Appendix 4*), which indicated *Tex19* was generally depleted of histone modifications and supports the hypothesis that DNA methylation is the primary and exclusive epigenetic mark mediating silencing of *Tex19* in non-expressing cells.

Notably, the relative enrichment of both H3K4me2 and H3K27me3 is altered in DP cells. This is consistent with DNA methylation and histone modifications functionally interacting (Vire et al., 2006; Cedar & Bergman, 2009). However, the observed chromatin modification changes in DP cells must occur downstream of promoter demethylation and therefore would be predicted to be a consequence of transcriptional activity. Of interest, the enrichment of H2A.Z increased in DP cells,

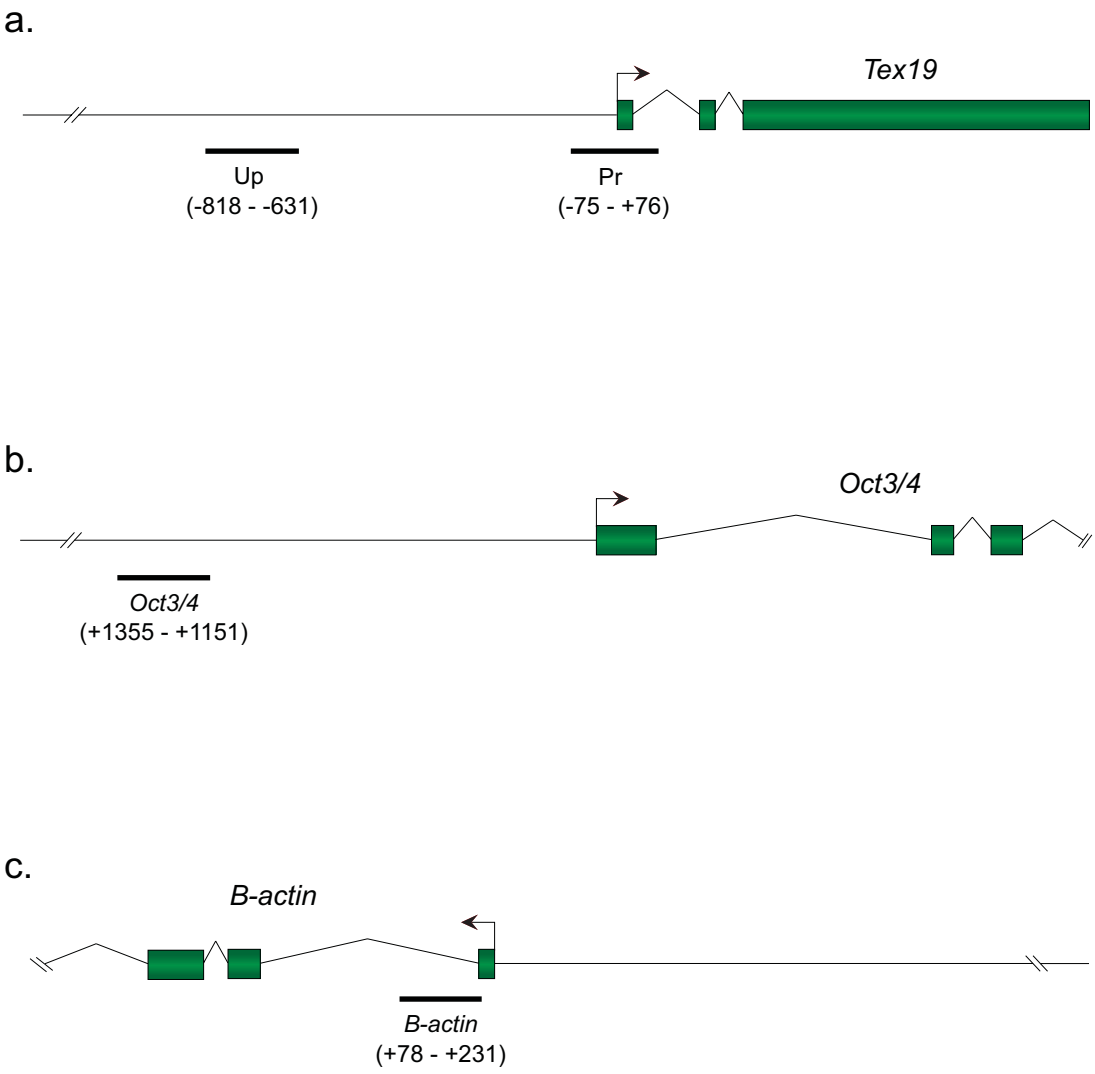


***Fig 4.11*. Schematic of genomic regions assayed by ChIP.** Green boxes represent the exon structure at each locus and arrows the transcriptional start site (TSS). Sites assayed by chromatin immunoprecipitation (ChIP) are shown as black boxes under the appropriate genomic region and the location given relative to the TSS. a.) Two regions of *Tex19* were assayed by ChIP. The promoter proximal region (Pr) and an upstream region (Up). b.) An upstream region of *Oct3/4* predicted to be enriched in H3K27me3 (Mikkelsen et al., 2008) was assayed as a positive control for this mark. c.) A promoter proximal region of *B-actin* predicted to be enriched in H3K4me (Mikkelsen et al., 2008) was selected as a positive control for this modification.

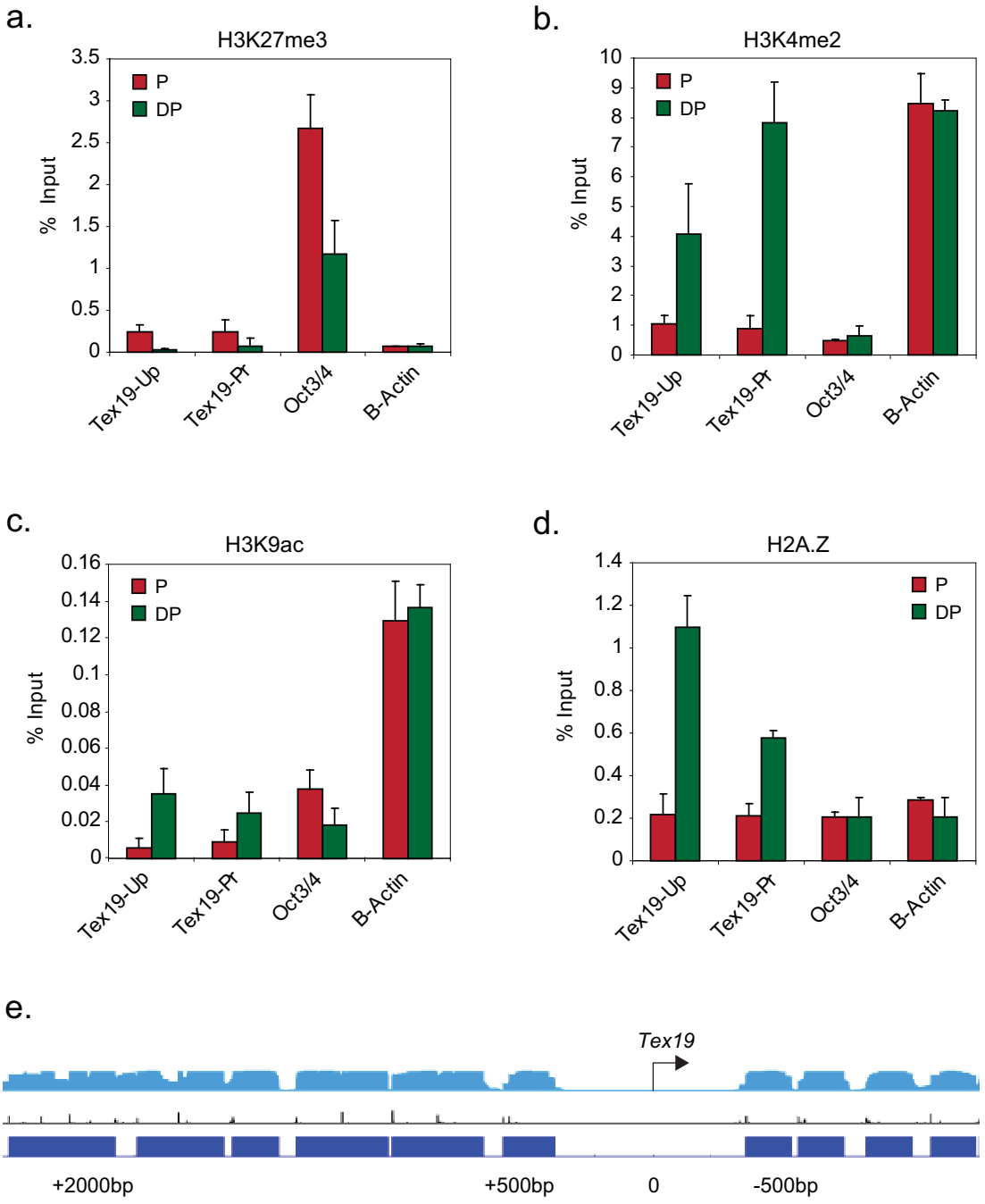


Fig 4.12. **ChIP analysis and nucleosome occupancy of *Tex19*.** a.) ChIP of H3K27me3 shows *Tex19* is depleted of this modification in P and DP cells. Importantly this mark is significantly depleted in non-expressing P cells indicating this modifciation is not a significant part of the repression mechanism at this locus. *Oct3/4* acts as positive control and *B-actin* represents background levels b.) *Tex19* is depleted of H3K4me2 in silenced DP cells but is enriched in hypomethylated DP cells. Here *B-actin* is the positive control and *Oct3/4* represents background levels c.) ChIP of H3K9ac and d.) histone variant H2A.Z in P cells and DP cells. e.) Computational prediction of nucleosome occupancy at *Tex19* using the Segal et al algorithm (2006). Each block represents the likely nucleosome positioning within the region. Note the absence of stable nucleosomes in the proximal promoter region.

consistent with reports from *Arabidopsis* that DNA methylation and H2A.Z deposition are mutually exclusive epigenetic states (Zilberman et al., 2008). In summary, this ChIP analysis strongly suggests that histone modifications are not significantly present at the *Tex19* promoter in repressed cell-types, consistent with promoter CpG methylation exclusively directing gene silencing.

Because neither activating nor repressive marks could be localised to the proximal *Tex19* promoter in nonexpressing cells I considered the possibility that this region was nucleosome-free. Extensive ChIP optimisation with the available H3 antibodies failed to enrich any loci above IgG controls, indicating that the antibody was ineffectual for this purpose. As an alternative, I used the computational algorithm devised by Segal et al (2006) (http://genie.weizmann.ac.il/software/nucleo_prediction.html) to determine nucleosome occupancy at *Tex19*. This analysis suggested that the promoter region (+500bp to -400bp) was strongly predicted to be devoid of stable nucleosomes (*Fig 4.12e*). Indeed the *Tex19* promoter sequence is significantly depleted of the periodic AA/TT dinucleotides that favour DNA bending and stable nucleosome assembly. Considered with the global histone modification profile (Mikkelsen et al., 2007) and the absence of chromatin modifications in the ChIP analysis performed here, this strongly suggests that the *Tex19* promoter could be refractory to nucleosome assembly and hence histone modifications. Therefore, this locus may critically rely on DNA methylation rather than histone modifications for epigenetic regulation. Ongoing collaborative experiments are being carried out here to confirm *Tex19* is nucleosome-free *in vitro* and *in vivo.* Interestingly, activation of nucleosome-free CGI promoters has been reported to be SWI/SNF independent, potentially suggesting why *Tex19* was unexpectedly resistant to demethylation in the absence of the putative SWI/SNF remodeller Lsh (Ramirez-Carrozzi et al., 2009)**.**

4.10 Tex19 remains silenced in differentiating ES cells lacking histone-modifying proteins

To functionally confirm that histone modifications do not have a crucial role in the regulation of *Tex19* I used ES cells with homozygous deletions for *Eed* or *Hdac1*

(Niswander et al., 1988; Zupkovitz et al., 2006). Eed is a critical component of the PRC2 complex and ES cells lacking this protein are unable to impose polycomb mediated silencing via the H3K27me3 modification (Montgomery et al., 2005; Boyer et al., 2006). Hdac1 is a histone deacetylase that contributes to several repressive complexes (Li j et al., 2002b; Zhang y et al., 1999). The absence of each of these proteins will enable the analysis of the role of the central polycomb and deacetylation repressive pathways at *Tex19.* I initially differentiated *Eed*-null, *Hdac1*-null or parental (wild-type) ES cells through embryoid body (EB) formation as previously. However, in the absence of Eed or Hdac1, ES cells failed to differentiate, as determined by *Oct3/4* expression and morphological criteria (data not shown). This suggests that Eed and Hdac1 are both critically necessary for cellular differentiation through EB formation. Therefore in this context, analysis of *Tex19* expression is uninformative.

To address this, I used an alternative experimental approach to induce ES cell differentiation. Here, I cultured *Eed*-null, *Hdac1*-null or parental (wild-type) ES cells with retinoic acid (RA) for three days. To validate the approach I also treated *Dnmt3a*-null, *Dnmt3b*-null and *Dnmt[3a 3b]*–null ES cells with RA for 3 days. Here qRT-PCR analysis demonstrated *Oct3/4* was strongly downregulated compared to mock-treated ES cells, indicating *Dnmt3* family-null ES cells undergo differentiation in response to RA (*Fig 4.13a & b*). Consistent with the EB studies, *Tex19* was silenced in RA treated WT and *Dnmt3a*-null ES cells but remained strongly expressed in *Dnmt3b*-null and *Dnmt[3a 3b]*–null ES cells (*Fig 4.13a & b*). This confirms the previous conclusion that Dnmt3b mediated *de novo* methylation is critically required for silencing *Tex19.*

To evaluate the role of Eed and Hdac1, I determined *Tex19* expression in mutant cells by qRT-PCR. After RA treatment both *Eed*-null and *Hdac1*-null ES cells repressed *Oct3/4* expression, indicating they have initiated differentiation in this approach (*Fig 4.12c & d*). Analysis of Eed-deficient ES cells showed *Tex19* was strongly silenced after 3 days RA treatment. This is consistent with *Tex19* repression occurring independently of H3K27 methylation and polycomb complexes. However

intriguingly, *Tex19* was not repressed in RA treated Hdac1-deficient ES cells (*Fig 4.13a & c*). This observation supports a critical role for Hdac1 and histone deacetylation in the initiation of *Tex19* silencing. However, treatment of somatic cells with the histone deacetylase inhibitor TSA, did not de-repress *Tex19* (*Fig 3.12a*). This suggests that deacetylation is insufficient to reactivate stably silenced *Tex19* but that it may have a role in initiating silencing. One possibility is that deacetylation is required to create a permissive environment or 'prime' *Tex19* chromatin for *de novo* methylation (Bird, 2002). However, this is not compatible with the putative conclusion that the *Tex19* promoter may be nucleosome-free.

It is therefore prudent to consider the alternative possibility that *Hdac1*-null cells have initiated atypical lineage-specific differentiation or have not differentiated at all here, as has been reported previously, and are thus are uninformative for this analysis (Lee et al., 2004). In this scenario *Oct3/4* could impose RA responsive silencing independent of Hdac1, but other differentiation programmes would not initiate and therefore *de novo* methylation activity (at *Tex19*) may not be induced. This notion is supported by the relatively moderate level of *Oct3/4* repression (compare Hdac1-/- with Eed-/-) and the fact that *Hdac1*-null cells were unable to differentiate into EBs. Indeed preliminary optimisation experiments with wild-type OS25 ES cells, which carry a *Oct3/4* selectable marker, suggested endogenous *Oct3/4* silencing in the absence of other differentiation events was a conceptual possibility. Here, 10 days EB differentiation of OS25 ES cells resulted in silencing of *Oct3/4* but not of *Tex19*. Crucially, bisulphite sequencing demonstrated that *Tex19* had not acquired any promoter methylation during this period suggesting the normal programme of differentiation had not occurred (*Tex19* is methylated in differentiated cells *in vivo*). This probably reflects the unique properties of OS25 cells as a result of extended *Oct3/4* selection (for example they appear morphologically fibroblast-like). Likewise Hdac1 deficient cells may have acquired unique plasticity properties as a result of global deacetylation. Thus, further studies are needed to determine the precise contribution of Hdac1 to *Tex19* silencing as it is not clear whether *Hdac1*-null ES cells differentiate here or simply repress *Oct3/4* in response to RA.

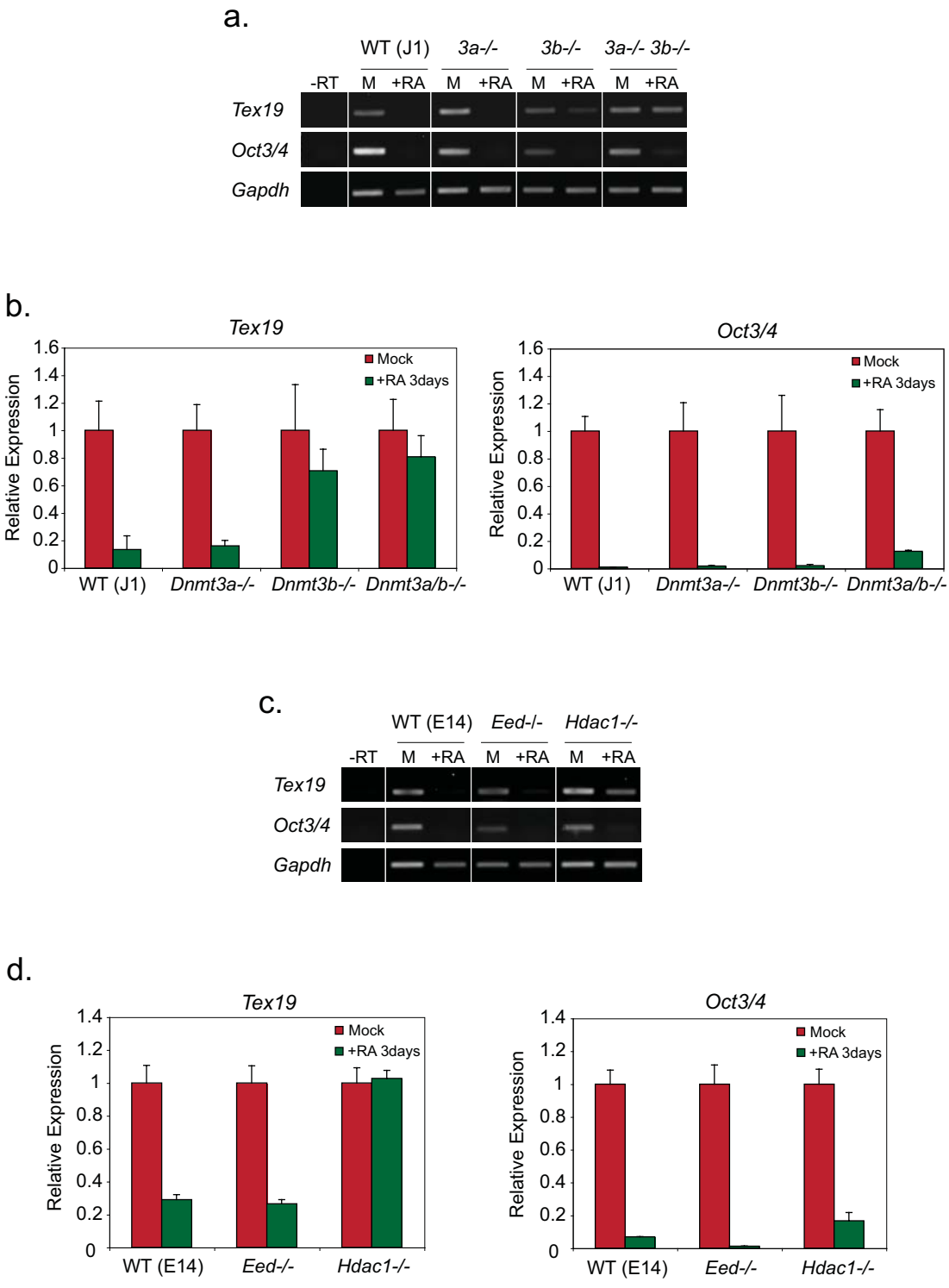


***Fig 4.13. Tex19* requires Dnmt3b and Hdac1 for silencing in RA differentiated ES cells.** a.) RT-PCR and b.) qRT-PCR analysis of *Tex19* expression in retinoic acid (RA) differentiated mutant ES cells. Consistent with EB studies Dnmt3b is required for silencing *Tex19* during RA differentiation. c.) RT-PCR and qRT-PCR analysis of *Eed*-null and *Hdac1*-null ES cells differentiated with RA. *Oct3/4* is repressed in all mutant ES cells indicating they are differentiating after RA addition. *Tex19* is silenced in differentiated *Eed*-null ES cells but not *Hdac1*-null ES cells.

In any case, these experiments have confirmed that *de novo* methylation is critically required for *Tex19* silencing and that polycomb histone modifications are entirely dispensable. Another possibility is that G9a, which catalyses H3K9 methylation, may play a role in repressing *Tex19.* Indeed G9a has been shown to target DNA methylation to several gene promoters (Tachibana et al., 20008; Dong et al., 2008). However, recent reports have indicated that G9a knockdown, leading to a global H3K9me2 reduction, has no effect on *Tex19* expression compared to controls (Chaturvedi et al., 2009: Yokochi et al., 2009 GEO accession database). Moreover in the absence of G9a, *Tex19* still acquires *de novo* methylation in differentiating ES cells (Epsztejn-Litman et al., 2008 supp data). In the present study, specific inhibition of G9a (and hence H3K9me2) with Bix-01294 (Kubicek et al., 2007) did not de-repress *Tex19* in pMEFs and the combination of Bix-01294 and TSA was also insufficient to de-repress *Tex19* (data not shown). In contrast 5-aza dC strongly activated expression (*Fig 3.11b*). Taken together these data strongly suggest that neither polycomb mediated H3K27me3 nor G9a mediated H3K9me2 have a significant role in regulating *Tex19.* Considered with the general depletion of histone modifications at *Tex19*, this provides compelling evidence that DNA methylation is the primary and exclusive mechanism regulating lineage specific *Tex19* expression.

4.11 DNA methylation regulates the lineage specific expression of Tex19 in PGCs

In the present study I have presented strong evidence that promoter methylation directs tissue-specific expression of *Tex19.* Because previous reports have suggested that methylation may regulate the activation of some germline specific genes during PGC development, I considered the possibility that promoter methylation also directs the temporal activation of *Tex19* (Maatouk et al., 2006). PGCs are specified in the extra-embryonic mesoderm at E7.25 and then commence actively migrating through the hindgut endoderm by E8.5 (Ginsberg et al., 1990). They colonise the gonads between E10.5-E11.5 and enter a phase of epigenetic reprogramming which includes global erasure of DNA methylation and histone modifications (Hajkova et al., 2002; Hajkova et al., 2008). During this time, the *Dazl* promoter is demethylated and this

correlates with the temporal activation of *Dazl* expression. To investigate whether *Tex19* undergoes a similar demethylation and activation phase in epigenetically reprogrammed PGCs, I FACS sorted migratory E10.5 PGCs and post-migratory E13.5 PGCs from embryonic gonads according to expression of a transgenic *Oct3/4* driven GFP marker (*Fig 4.14a*) (Boiani et al., 2002). This process resulted in populations of PGCs that were typically >95% (E10.5) or >98% (E13.5) pure, as judged by staining for endogenous Oct3/4 (*Fig 4.14b*) and Nanog (*Appendix 5*). In addition, I was unable to detect expression of the somatic cell marker *Sf1* in purified PGCs by qRT-PCR but could readily detect expression in the remaining somatic cells of the sorted gonad (*Fig 4.14c*). I was also able to confirm that visual discrimination between male and female gonads at E13.5 was accurate, as male E13.5 PGCs exclusively expressed *Dnmt3l,* as expected (*Fig 4.14c*) (Small et al., 2005 supp data).

To validate the approach I examined the methylation status of *Dazl* in E10.5 and E13.5 male and female PGCs. Consistent with previous reports, *Dazl* was found to be methylated in E10.5 PGCs but hypomethylated by E13.5 (Maatouk et al., 2006). In contrast, the somatic cells of the gonad remained methylated throughout this period. Interestingly, while previous reports have noted the correlation between *Dazl* demethylation and upregulation of protein levels in PGCs, this has not been confirmed to be a transcriptional effect. To assess this, I performed qRT-PCR on sorted PGC populations. An important aspect of qRT-PCR analysis of developing PGCs is selection of an appropriate reference (normalising) gene, particularly with limited cell numbers. This is because the global transcriptional profile of PGCs alters dramatically between E10.5 and E13.5, including expression of commonly used reference genes. To identify a suitable reference gene, I tested five candidates, *Oct3/4, Atp5b, Gapdh, Hprt* and *18S rRNA* and used the geNorm algorithm to evaluate the most stably expressed gene (Vandesompele et al., 2002). This analysis identified 18S rRNA as maintaining constant expression between samples and also at a similar level to the algorithm derived average of all 5 candidates. 18S rRNA has also been used as a reference gene in several previous PGC studies (Svingen et al., 2007; Beverdam & Koopman, 2006; Silva et al., 2008). Using 18S rRNA to

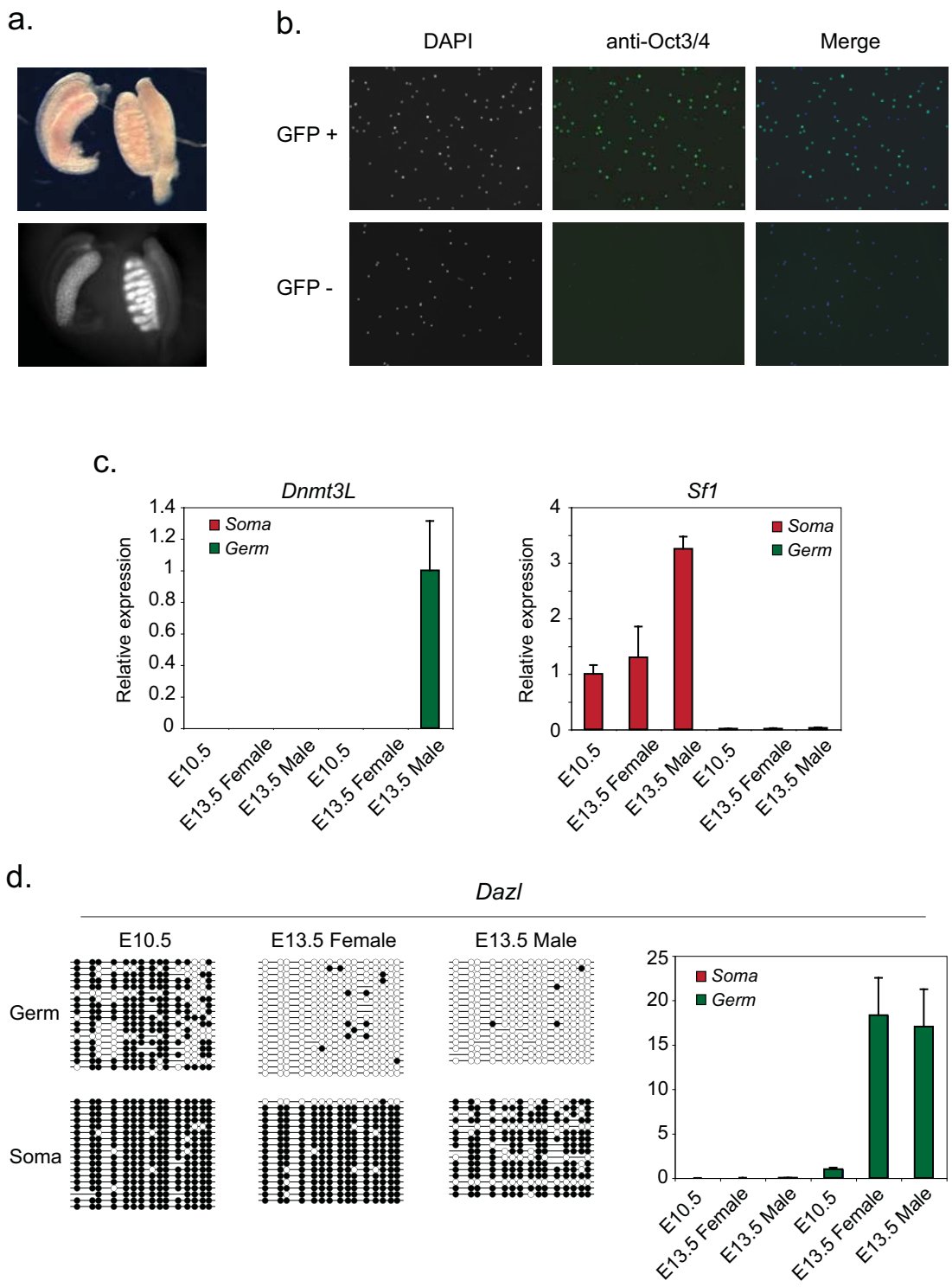


Fig 4.14. **Quality control analysis of GFP sorted primordial germ cells and somatic cells.** a.) Male (right) and female (left) gonads with PGCs expressing transgenic Oct3/4 GFP were dissected from E10.5 or E13.5 (shown) embryos. b.) Staining for endogenous Oct3/4 from FACS sorted cells. Shown is E13.5 males cells which were typically >98% pure. c.) qRT-PCR analysis demonstrating male sorted PGCs express the male specific marker *Dnmt3l* (left panel) and that only somatic cells express the somatic gonad marker *Sf1* (right panel). d.) Methylation analysis of *Dazl* showing the promoter is methylated in E10.5 migratory PGCs but demthylated by E13.5. This correllates with increased expression of *Dazl* in E13.5 PGCs by qRT-PCR (right panal). Expression is relative to E10.5 germ cells, which are set to 1 and normalised to 18S RNA expression.

normalise expression, qRT-PCR showed that demethylation of *Dazl* does correlate with strongly increased transcription after epigenetic reprogramming (*Fig 4.14d*). These results suggest I have generated purified populations of PGCs that are consistent with expression and methylation analyses from previous studies.

To test whether *Tex19* showed similar promoter dynamics to *Dazl* during PGC maturation I investigated *Tex19* methylation in E10.5 and E13.5 PGCs. Surprisingly, I found *Tex19* to be hypomethylated prior to epigenetic reprogramming at E10.5 in addition to hypomethylation of male and female PGCs at E13.5 (*Fig 4.15a*). As expected, *Tex19* was heavily methylated in the somatic cells of the gonad. Notably, qRT-PCR indicated there was no significant change in *Tex19* expression between E10.5 and E13.5 PGCs (*Fig 4.15a*). This data suggests that, unlike *Dazl, Tex19* is unmethylated and expressed in PGCs prior to entry into the gonad and epigenetic reprogramming. To ensure that *Tex19* had not entered early epigenetic reprogramming in E10.5 PGCs, I examined the methylation of sorted E9.5 PGCs. Here, *Tex19* was also found to be severely hypomethylated (*Appendix 5*), strongly supporting the hypothesis that *Tex19* is demethylated prior to entry into the gonads. Analysis of *PiwiL2* promoter dynamics and expression indicated that this gene followed a similar pattern to *Tex19* and was demethylated and expressed prior to epigenetic reprogramming (*Fig 4.15b*). In contrast *Tex19.2* was methylated in E10.5 PGCs and demethylation in E13.5 male and female PGCs correlated with transcriptional activation (*Fig 4.15c*). These data indicate that promoter methylation does not regulate the temporal activation of *Tex19* at ~E11.5 during the PGC-specific epigenetic reprogramming event.

Interestingly, expression analysis of several other germline specific genes demonstrated that there is a bimodal distribution of expression at this stage, such that genes are either strongly activated between E10.5 and E13.5 or exhibit no significant expression change (*Fig 4.16a* & appendix for qRT-PCR). Based on methylation analysis of *Tex19, PiwiL2* and *Tex19.2* and previous studies, I would predict that the genes expressed prior to entry into the gonad would be previously hypomethylated whereas the genes activated at E13.5 undergo promoter demethylation coincident

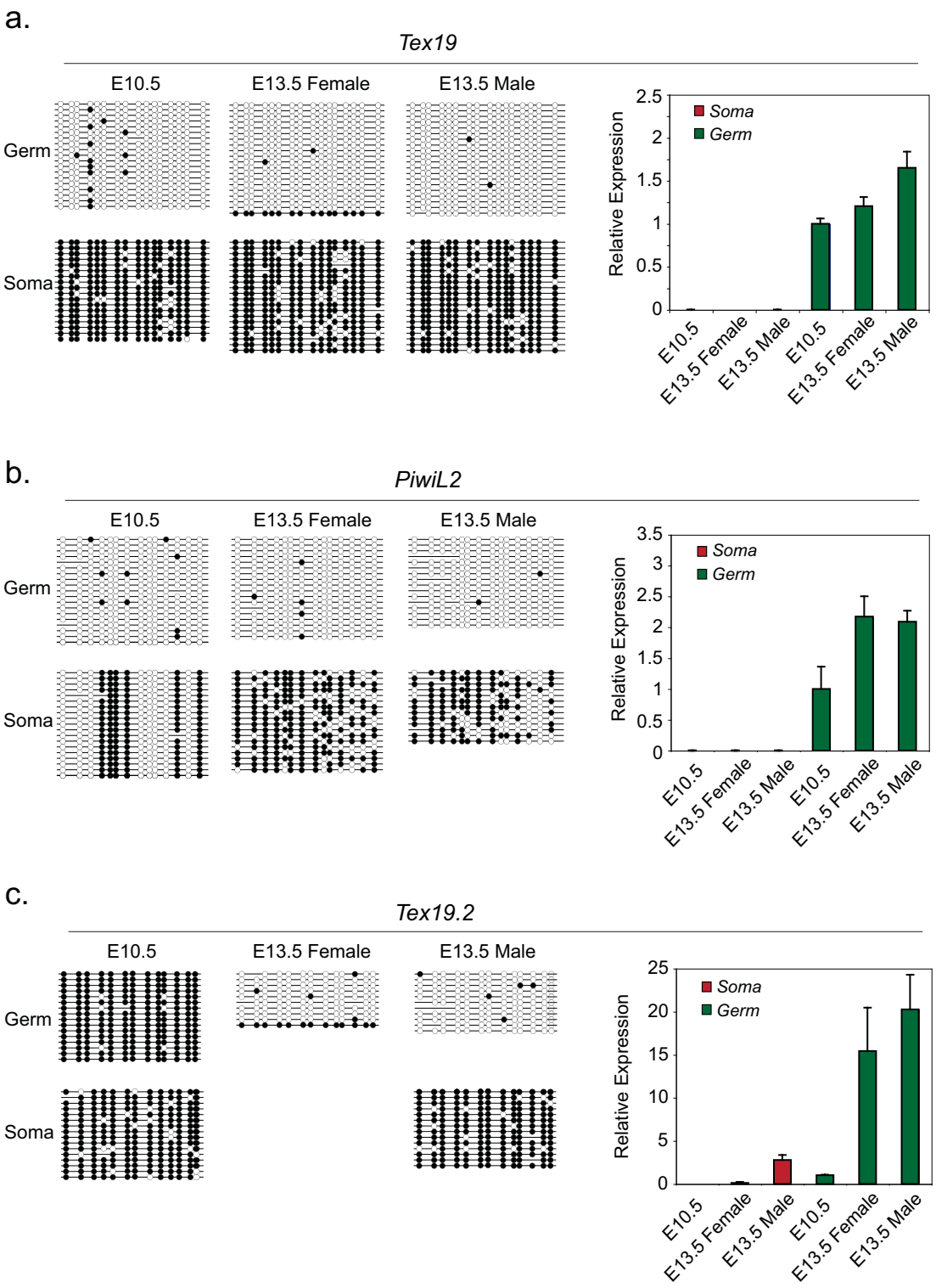


Fig 4.15. **Analysis of methylation status and expression in GFP sorted E10.5 and E13.5 primordial germ cells.** a.) Promoter methylation of *Tex19* (BS1) in E10.5 PGCs, E13.5 male or female PGCs and matched somatic cells of the gonad. Right panel, qRT-PCR analysis of *Tex19* expression. b.) Methylation and expression of *PiwiL2* from somatic cells and PGCs of E10.5 embryos and male and female E13.5 embryos. c.) *Tex19.2* methylation and expression analysis. Note missing panel for somatic E13.5 females. Successive attempts to PCR this sample failed, however, it would be strongly predicted that *Tex19.2* would be methylated in E13.5 female somatic gonad tissue. Expression is relative to E10.5 PGCs, which is set to 1. and normalised to 18S RNA. Error bars represent the propergated S.E.M.

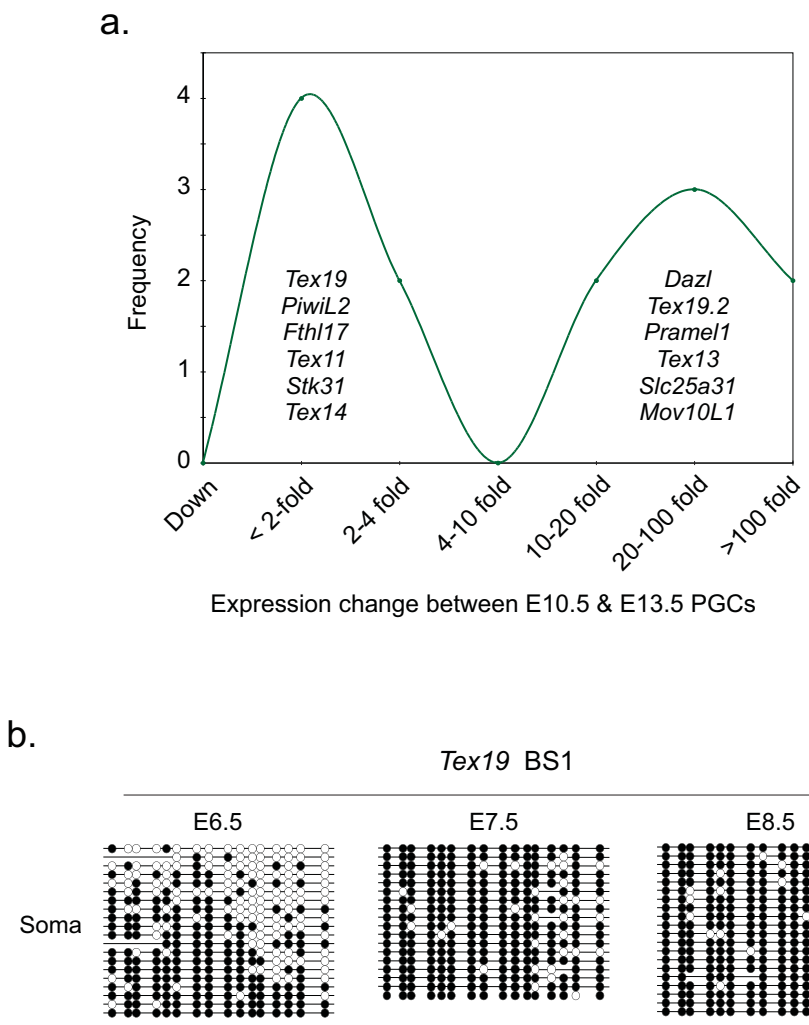


***Fig 4.16. Tex19* is methylated in somatic cells by E7.5.** a.) Germline genes tested fall into one of two distinct categories depending on expression change between E10.5 and E13.5. The first set (left peak) are expressed prior to entry into the gonad at ~E11.5 and therefore exhibit limited or no change in expression. These loci, like *Tex19*, would be predicted to be hypomethylated at E10.5. The second set (right peak) are strongly upregulated by E13.5 and like *Dazl*, would be predicted to be hypermethylated at E10.5 and demethylated coincident with expression by E13.5. (see *Appendix 5*) b.) *Tex19* is fully methylated by E7.5 in whole embryonic tissues coincident with gene silencing throughout the embryo. This suggests early PGCs could specifically demethylate and activate *Tex19*.

with epigenetic erasure. Notably, several of the genes that fall into this latter category (*Tex13, Tex19.2, Mov10l1*) were included in the candidate gene list generated in *Chapter 3*. These genes may represent novel targets crucially reliant on DNA methylation for lineage-specific expression and temporal activation in ~E13.5 post-migratory PGCs similarly to *Dazl* (*Appendix 5*).

Because *Tex19* is hypomethylated in migratory E10.5 PGCs but methylated in somatic tissues, I investigated the developmental time-point at which lineage-specific methylation occurs. *Tex19* is demethylated and expressed in the zygote and ES cells, therefore there were two broad possibilities. Firstly, *Tex19* becomes methylated in only somatic cells *after* PGC specification (after ~E7.5), which themselves avoid *de novo* methylation or secondly, *Tex19* is methylated in all lineages during zygotic remethylation (between E4.5-E6.5) and is then specifically demethylated in early PGCs. To distinguish between these possibilities I examined the methylation of *Tex19* in the embryonic tissues of E6.5 to E8.5 embryos. Here the *Tex19* promoter was found to be methylated at E7.5 and E8.5 but interestingly, in an intermediate state in E6.5 embryos. (*Fig 4.16b*) This suggests that *Tex19* is undergoing a process of *de novo* methylation until E6.5 that is complete by E7.5. Notably, *Tex19* protein is expressed throughout all lineages from the zygote stage until late E7.5 embryos (personal communication - *J. Reichmann*). This data supports the second hypothesis that *Tex19* is methylated in all lineages and is specifically demethylated in early PGCs. Moreover, the promoter methylation data is consistent with expression of *Tex19* during embryonic development being regulated by CpG methylation, probably mediated by Dnmt3b, which is strongly expressed at this time. Importantly, similarly to EB formation (*Section 4.6*), *Tex19* acquires DNA methylation prior to repression suggesting promoter methylation is the upstream and primary mechanism of silencing.

PGCs have been reported to enter a preliminary phase of epigenetic reprogramming at ~E8.5 including partial erasure of DNA methylation (Seki et al., 2007; Hajkova et al., 2008). A likely possibility is that *Tex19* is specifically demethylated in PGCs during this event leading to germline specific expression of this transcript. Consistent

with this, Tex19 protein has been detected in PGCs at E9.5 (Kuntz et al., 2008). Based on the experimental evidence presented here, *Tex19* represents a novel gene regulated by lineage-specific demethylation in PGCs before E9.5. Furthermore, *Tex19* is the first reported gene to unequivocally and exclusively rely on promoter methylation for directing its *in vivo* expression pattern and epigenetic memory independently of other epigenetic and trans-acting factors.

4.12 Discussion

The precise role of DNA methylation in regulating *in vivo* patterns of gene expression is currently unclear. Genetic studies have demonstrated that loss of methylation leads to global gene de-repression (Jackson-Grusby et al., 2001; Lande-Dinder et al., 2006; Fouse et al., 2008) and that developmental or tissue specific expression of some genes correlates with promoter methylation suggesting methylation may regulate some loci (Maatouk et al., 2006; de Smet et al., 1999; Rodic et al., 2005). However crucially, these reports do not demonstrate a cause and effect relationship. Therefore there are currently no examples of genes that crucially and exclusively rely on promoter CpG methylation for regulation in all lineage contexts. This chapter has progressively demonstrated that transcription factor availability and histone modifications do not have a significant role in regulating expression of the germline restricted gene *Tex19.* Instead, promoter CpG methylation is the primary and exclusive mechanism of directing developmental lineage-specific *Tex19* expression.

The CpG-dense *Tex19* promoter is heavily methylated in silenced tissues but hypomethylated in expressing cells *in vivo.* Furthermore, *de novo* promoter methylation is targeted to *Tex19* upstream of gene silencing during ES cell differentiation. Genetic inhibition of *de novo* methylation by *Dnmt[3a 3b]* deletion prevents *Tex19* silencing in ES cells but not *Oct3/4* silencing, strongly supporting methylation as the key event directing regulation of *Tex19.* This key observation demonstrates the cause and effect relationship between promoter CpG methylation and *Tex19* silencing. Consistent with this, promoter methylation strongly repressed

expression of a *Tex19* reporter. Moreover, experimentally induced demethylation by 5-aza dC treatment or Dnmt1 deletion strongly activates *Tex19* expression in somatic cells. In contrast treatment with TSA or Bix-01294 had no effect. Following demethylation and gene activation, transcriptional silencing is unable to be re-imposed upon *Tex19* suggesting DNA methylation is the critical component directing epigenetic memory of the gene. *In vivo*, *Tex19* is likely *de novo* methylated by Dnmt3b between implantation and E6.5-E7.5, after which expression is silenced in somatic cells. PGCs avoid methylation or are specifically demethylated, leading to lineage-specific expression of *Tex19* in germ cells from at least E9.5. According to this study and others, the core *Tex19* promoter is depleted of histone modifications, which may reflect the inherent instability of nucleosome-assembly at *Tex19*. Furthermore, the proximal promoter of *Tex19* can drive precocious expression in all cell-lineages investigated, suggesting an epigenetic mechanism silences endogenous *Tex19* in somatic cells. Taken together, these observations strongly support promoter CpG methylation as the primary and upstream mechanism of regulating *in vivo* expression of *Tex19.* To my knowledge, this is the first cause and effect report of a gene directly and exclusively regulated by developmental and tissue-specific DNA methylation.

This study also revealed that promoter CpG methylation is an important component of the regulatory system of *PiwiL2.* Here, in the absence of *de novo* methylation complete silencing cannot be imposed on *PiwiL2* during ES cell differentiation. Moreover *PiwiL2* is specifically hypomethylated in expressing tissues and demethylation by 5-aza dC or Dnmt1 deletion strongly activates gene expression in somatic cells. The *PiwiL2* promoter can also drive moderate expression (although significantly lower than *Tex19*) in somatic cell types indicating general transcription factors can activate the gene, which therefore does not rely on germline specific factors. However crucially, in this study *PiwiL2* was significantly repressed, but not silenced, in differentiating *Dnmt[3a 3b]*-null ES cells. This indicates that while DNA methylation may be an important and crucial contributor to *PiwiL2* regulation, other epigenetic or trans-acting systems must also target the locus. Notably, the other CGI candidate loci tested here were also specifically hypomethylated in expressing

germline tissues. It is possible that these genes could exclusively rely on promoter methylation for repression *in vivo*, similarly to *Tex19*. Intriguingly, many of these genes (*Mov10l1, Tex13, Tex19.2)* are strongly activated during PGC epigenetic reprogramming at ~E11.5. It is therefore possible that these genes represent novel targets that utilize the PGC-specific wave of demethylation at ~E11.5 to mediate their temporal and germline specific activation.

Chapter 5

Kaiso mediates methylation-dependent silencing of *Tex19*

5.1 Introduction

DNA methylation was recognized as a mechanism that stably silences transcription over 30 years ago (Holliday & Pugh, 1975; Razin & Riggs 1980; Felsenfeld & McGhee, 1982;). However, the precise mechanism through which CpG methylation interferes with transcriptional activation still remains undetermined. Two models have been proposed, both of which could be biologically relevant. Firstly, methylated-CpGs may directly exclude regulatory proteins from binding to their target DNA sites (Kovesdi, et al. 1987). Secondly, methyl-CpG binding proteins (MBPs) may be recruited to methylated target sites where they direct transcriptional repression (Meehan., et al 1989; Boyes & Bird, 1991; Bird, 2002). The former model is supported by studies of the chromatin insulator protein CTCF, which binds the unmethylated maternal *Igf2* enhancer, but is inhibited from binding the paternal copy due to dense CpG methylation (Hark et al., 2000; Szabo et al., 2000). Additionally, methylated target sequences interfere with the binding of CREB and E2F transcription factors (Iguchi-Ariga & Schaffner, 1989; Campanero et al., 2000). However, while these examples provide strong evidence for the biological relevance of the direct exclusion model, further reports of DNA methylation inhibiting protein binding *in vivo* have not emerged and indeed, some factors have been demonstrated to be impervious to CpG methylation within their binding site (Zhu et al., 2003). Furthermore many important DNA-binding proteins lack a CpG dinucleotide in their consensus recognition motif, rendering them insensitive to DNA methylation. These observations suggest that the direct exclusion model is not applicable as a general methylation-dependent silencing mechanism. Thus, the alternative mechanism

whereby MBPs are selectively targeted to methylated CpGs, has emerged as the favoured paradigm for methyl-dependent gene repression. In this model, complexes containing MBD or Kaiso family proteins bind methyl-CpGs via their MBD or zinc-finger domains, respectively, and target co-repressor activity to bound loci (*Section 1.4.6.1*) (Sasai & Defossez, 2009).

The recruitment of methyl-binding repressor proteins to methylated target sites provides an attractive mechanism through which transcriptional silencing can be targeted to methylated domains. However, similarly to the direct exclusion model, the recruitment of MBPs fails to explain all the experimental evidence and predictions expected of a global methylation-dependent silencing mechanism. Here, the early *in vitro* evidence demonstrating that MBPs repress methylated reporters has not been supported by recent genetic and *in vivo* studies (Boyes & Bird, 1991; Martin Caballero et al., 2009). One substantial area of contention are the genetic knockouts for each MBD or Kaiso gene. These mutant mice are viable and exhibit no overt phenotype, suggesting methyl-binding proteins do not have a general role in regulating developmental expression patterns (Hendrich et al., 2001; Zhao et al., 2003; Prokhortchouk et al., 2006; Chen et al., 2001; Guy et al., 2001). Indeed the recently generated *MeCP2*, *MBD2* and *Kaiso* triple-null mice are also viable demonstrating that these three proteins do not functionally compensate for each other in this respect (Martin Caballero et al., 2009). Furthermore, analysis of mutant MeCP2 or MBD2 cells revealed that only modest changes in gene expression occur (Tudor et al., 2002; Hutchins et al., 2002; Kriaucionis et al., 2006 Berger et al., 2007) although a significant role for MBD2 has been noted in contributing to *Xist* repression (Barr et al., 2007). Additionally, while there is some evidence that Kaiso may target the *MTA2* gene in aberrantly methylated HeLa cells, currently there are also no examples of methyl-dependent genes regulated by Kaiso in normal cells (Yoon et al., 2003; Sasai & Defossez, 2009).

One possibility for the mild phenotype and the lack of gene expression changes in MBP deficient cells, is that remaining MBPs (those not knocked-out in mice as of yet) are functionally degenerate and compensate for loss of other MBPs, a notion that

is supported by some genetic evidence (Martin Caballero et al., 2009). Alternatively, DNA methylation may inherently repress transcription through one or a combination of other mechanisms (such as transcription factor exclusion, Dnmt co-repressor recruitment or chromatin condensation) and thus methyl-binding proteins may have subtle roles in fine-tuning gene silencing. In any case, despite the current favoured model predicting that MBPs are the primary mediators of methylation-dependent repression, the current body of evidence offers little support for this model and convincing examples of genes clearly regulated by a methyl-binding protein *in vivo* are rare in the literature. Thus, the precise mechanism through which DNA methylation silences transcription on a global scale remains unresolved.

In the present study I have identified a gene, *Tex19,* which relies exclusively on promoter CpG methylation for its tissue-specific and developmental expression pattern. In this chapter, I consider whether methyl-binding proteins could be the mechanism of coupling *Tex19* methylation to gene silencing. I reasoned that for a gene to critically rely on MBPs for repression it must also critically depend on CpG methylation. Therefore, *Tex19* represents an excellent candidate for MBP-mediated repression. Moreover, defining the mechanism of methyl-dependent silencing would provide compelling support for the argument that *Tex19* is primarily and exclusively regulated by DNA methylation. The analysis presented here strongly suggests that Kaiso is a primary mediator of silencing at *Tex19*, and that *Tex19* therefore represents a novel example of a methylation-dependent gene regulated by an MBP in normal cells. In this chapter, I also investigate the potential mechanisms that could specifically target the *Tex19* CGI promoter for *de novo* methylation during development.

5.2 *Tex19* is de-repressed in Kaiso-null fibroblasts

As part of a strategy to determine the role of MBPs at *Tex19,* I initially examined *Tex19* expression in single mutant *MBD2*$^{-/-}$ or *Kaiso*$^{-/-}$ mouse fibroblasts and *MeCP2*, *MBD2* and *Kaiso* triple-null (MKO) fibroblasts. These simian virus-40 transformed (SV40) cell lines were derived from the tails of mutant or wild-type mice (Barr et al.,

2006). RT-PCR genotyping of each cell line revealed that the *MBD2*$^{-/-}$ and *Kaiso*$^{-/-}$ fibroblasts were genotyped as predicted (*Fig 5.1a*). However, while MKO fibroblasts did not express *MBD2* and *MeCP2*, a weak band of *Kaiso* expression could be detected (*Fig 5.1a*). qRT-PCR showed that this Kaiso expression was in the order of ~30-fold lower than parental wild-type fibroblasts (data not shown). It is unclear whether there is a contaminating sub-population of double-null cells (*MeCP2* & *MBD2*) among the MKO population or residual Kaiso is expressed at low levels in all MKO cells. Because Kaiso transcripts were not detected at any level in the single *Kaiso*-null cells, the former possibility is most likely here. Importantly, in any case, the low levels of Kaiso present (~3%) would be predicted not to maintain significant biological activity or significantly affect population expression analysis, indicating these cells can be considered as *de facto* MKO fibroblasts.

Unexpectedly, RT-PCR analysis of mutant MBP cells showed that *Tex19* was strongly de-repressed in *Kaiso*-null fibroblasts and reciprocally, in MKO fibroblasts. Limited *Tex19* transcript was also detected in *MBD2*-null cells (*Fig 5.1b*). This data suggests that Kaiso is a critical mediator of *Tex19* repression in somatic fibroblasts and that MBD2 may have a minor or complimentary role. To determine the precise degree of de-repression in MBP mutant fibroblasts I performed qRT-PCR. *Tex19* was found to be de-repressed ~45-fold in *Kaiso*$^{-/-}$ cells and ~39-fold in MKO cells relative to wild-type parental fibroblasts. The moderately lower expression levels in MKO fibroblasts relative to *Kaiso*-null cells was not significant (*students t-test p* = 0.23) and may reflect residual Kaiso present within the MKO population. The fact that MKO fibroblasts exhibited statistically equivalent expression of *Tex19* to *Kaiso*-null fibroblasts suggests that the weak activation observed in the *MBD2*-null fibroblasts is not additive with Kaiso-deficiency. This analysis thus indicates that Kaiso may be a primary mediator of methylation-dependent *Tex19* silencing

Because *Tex19* relies on promoter methylation to maintain epigenetic silencing I investigated the possibility that de-repression in MBP mutant fibroblasts was due to partial or complete demethylation of the promoter. Bisulphite analysis revealed that all MBP mutant cell lines examined retained fully hypermethylated *Tex19* relative to

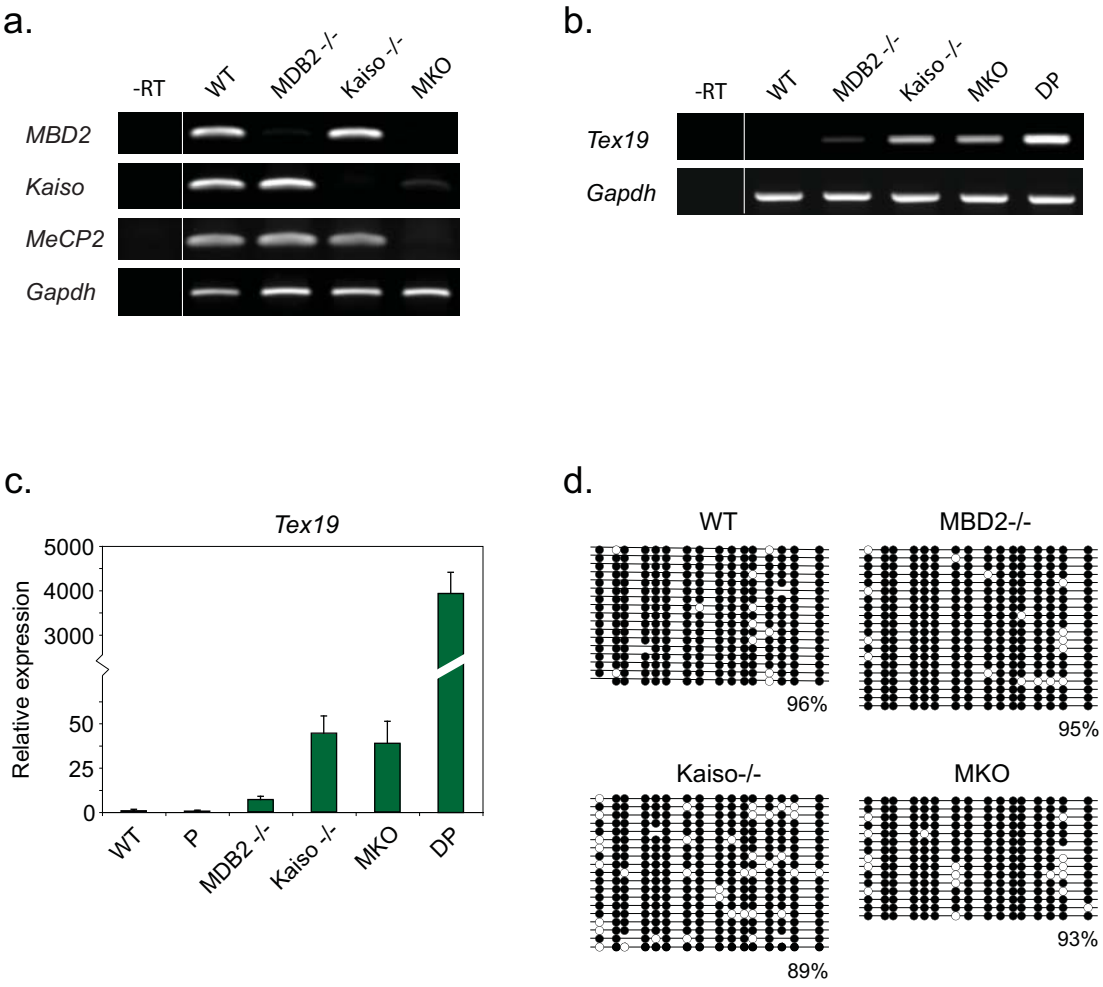


***Fig 5.1. Tex19* is expressed in *Kaiso-/-* fibroblasts but remains hypermethylated.** a.) RT-PCR genotyping of methyl-binding protein (MBP) deficient tail-tip fibroblasts. Cell genotypes are shown across the top panel. MKO are triple-null for *MBD2*, *Kaiso* and *MeCP2*. b.) RT-PCR and c.) qRT-PCR expression analysis of *Tex19* in MBP mutant cells. *Gapdh* was used to normalise loading. d.) *Tex19* BS1 promoter methylation in mutant MBP fibroblasts. *Tex19* was found to be methylated despite being strongly expressed in Kaiso-null and MKO cells.

wild-type fibroblasts and *in vivo* tissues (*Fig 5.1d* & *Fig 4.2c*). This data indicates that despite maintaining methylation of a CpG dense (HCP) promoter, strong transcription initiates from the *Tex19* locus in Kaiso-deficient cells. To my knowledge, *Tex19* represents a novel example in the literature of a heavily methylated CpG-dense gene being strongly transcribed above background (29-cycles here). This suggests that Kaiso represents a primary mechanism through which DNA methylation mediates transcriptional repression at *Tex19.* Interestingly, while strong expression was observed in Kaiso-null and MKO cells, it was significantly less than transcription levels in DP cells or wild-type testis (*Fig 5.1c*). This indicates that while Kaiso is a critical component of silencing at *Tex19,* it is not the only pathway through which methylation-dependent repression is imposed at the locus. One possibility is that *Tex19* promoter methylation *per se* directly inhibits transcription or that it excludes binding factors. Therefore, abrogation of Kaiso may only cause partial de-repression relative to hypomethylation. In summary, deletion of Kaiso and to an extent MBD2, leads to partial but significant de-repression of the methyl-dependent gene *Tex19* without changes in promoter methylation. This data is consistent with these MBPs being crucially required to impose full gene silencing at *Tex19* but also indicate that DNA methylation at this locus imposes partial repression through alternative mechanisms to the MBPs tested here.

5.3 Global analysis identifies a limited number of de-repressed genes in Kaiso-null fibroblasts

Because *Tex19* was unexpectedly de-repressed in MBP mutant fibroblasts, I considered whether any other candidate or germline-specific genes were activated in MBP-null cells. I investigated expression of germline transcripts by RT-PCR analysis. In contrast to *Tex19,* I could not detect de-repression of any genes tested in *Kaiso*-null fibroblasts (*Fig 5.2a*). This indicates that in the context of germline-specific genes, *Tex19* is a specific target of Kaiso*,* and that recruiting Kaiso is not a general mechanism of repressing methylated germline-specific genes. However intriguingly, *PiwiL2* was specifically de-repressed in *MBD2*-null and MKO fibroblasts. No other transcripts were found to be activated in *MBD2*$^{-/-}$ or MKO cells

(*Fig 5.2a*). qRT-PCR analysis suggested that *PiwiL2 was* de-repressed 4.2-fold in the absence of MBD2 but no change was detected in *Kaiso*-null cells (data not shown). This indicates that MBD2 could be mediating at least part of the methyl-dependent silencing of *PiwiL2* in somatic cells.

RT-PCR analysis suggested that MBP-mediated repression was not a broad mechanism for silencing methylated germline-specific genes but instead was restricted to specific but novel loci (*Tex19* and *PiwiL2*). To examine whether Kaiso had a role in global gene silencing, I investigated the genome-wide expression profile of Kaiso-null and MKO fibroblasts relative to wild-type using Illumina Ref-6 BeadChips. Consistent with the mild phenotype of Kaiso-knockout mice (Prokhortchouk et al., 2006), I found only 31 transcripts upregulated in Kaiso-null fibroblasts using a stringent 6-fold change threshold (*Fig 5.2b*) (*Table 5.1*). Notably, I was unable to detect expression changes in putative Kaiso target genes (*S100A4, MTA, Wnt11* & *Rapsyn*), in agreement with recent studies that indicated these genes are not regulated by Kaiso (Prokhortchouk et al., 2006; Ruzov et al., 2009a) However, I did detect strong de-repression of *Tex19* (>34-fold) in Kaiso-null cells. This supports the RT-PCR analysis suggesting *Tex19* silencing is dependent on Kaiso (*Fig 5.1b*). In contrast to Kaiso-deficient cells, 234 transcripts were upregulated in MKO cells. This is surprising given the mild-phenotype of triple-null mice (Martin Caballero et al., 2009). However, consistent with my array analysis, a study using siRNAs to specifically knockdown MBD2, MeCP2 and MBD1 in HeLa cells, reported that 15% of transcripts were significantly upregulated (Lopez-Serra et al., 2008). This suggests that abrogation of multiple MBPs can lead to de-repression of a significant number of genes but that the phenotypic consequences of this altered gene expression are relatively mild (Martin-Caballero et al., 2009). An alternative possibility is that the *ex vivo* culture and/or SV40 transformation of the mutant and wild-type fibroblasts used here (and HeLa cells elsewhere) has influenced expression patterns. However, the relatively modest number of expression changes in co-cultured *Kaiso*-null fibroblasts argues against this.

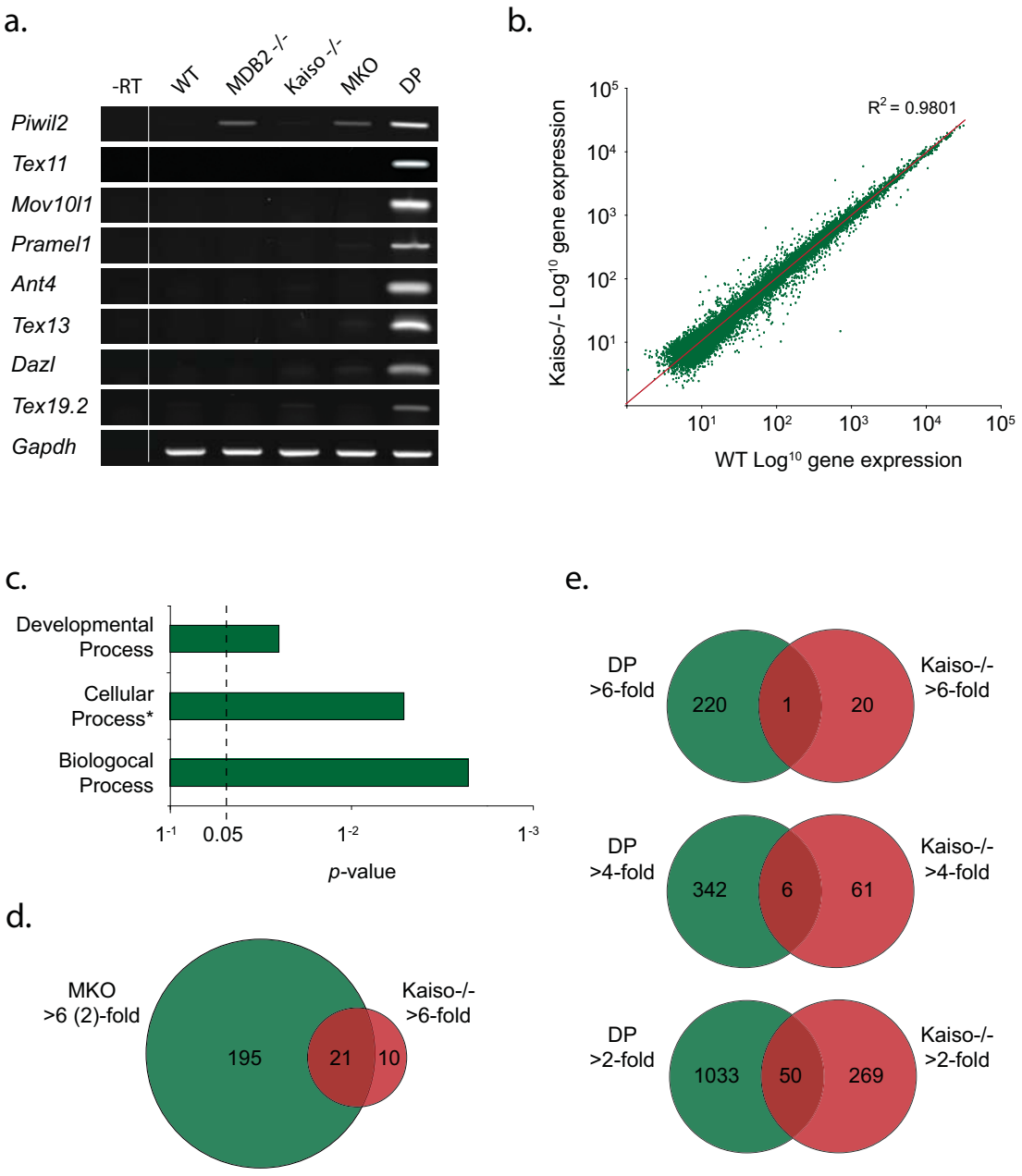


Fig 5.2. **No global gene de-repression of in Kaiso-null fibroblasts.** a.) RT-PCR expression analysis of candidate and germline-specific genes in MBP mutant fibroblasts reveals only *PiwiL2 is* de-repressed. b.) Scatter plot of log[10] expression values from microarray analysis of global gene expression in wild-type and Kaiso-null fibroblasts. Few significant changes in expression were detected. For comparison see *Fig 3.6a*. c.) Gene ontology analysis of transcripts de-repressed >6-fold in Kaiso-null cells at hierarchial level BP1. d.) Scaled venn diagram cross-referencing genes de-repressed in Kaiso and MKO fibroblasts. To account for variations in de-repression, MKO genes are considered to overlap the >6-fold Kaiso dataset if they were de-repressed >2-fold. e.) Venn diagrams cross-referencing genes upregulated the indicated degree in DP MEFs and Kaiso-null fibroblasts. Genes de-repressed in both data sets would be predicted to be putitative methylation-dependent Kaiso targets. Note only *Tex19* is common between both dataset at the most stringent >6-fold threshold. Additional probes present in the Kaiso-\- microarray have been removed from this analysis. Not to scale.

Transcript	Chromosome	Fold-change Kaiso-/-	Fold-change MKO
Zic1	9	65.2	(1)
Tex19	11	34.5	21.2
Bcl11b	12	24.6	(1)
Mid1	X	19.6	3.5
Ebf3	7	18.7	(1)
Tssc8	7	11.6	2.8
Pcsk9	4	10.7	(1)
Erdr1	ND	10.2	10.3
LOC674912	ND	9.6	12.2
C330015L04Rik	ND	8.9	5.3
Igfbp2	1	8.7	52.4
Uhmk1	1	8.3	5.9
Frmpd4	X	7.7	4.7
Hic1	11	7.5	3.9
A630064P09Rik	ND	7.5	5.0
LOC100044257	ND	7.4	(1)
Lhx9	1	7.3	21.5
Igfbp5	1	7.2	-4.8
Megf6	4	7.2	(1)
1500009L16Rik	10	7.0	9.0
LOC100039175	14	6.9	0.9
Mgp	6	6.9	3.0
Klhl10	11	6.8	1.5
Gdf1	8	6.7	(1)
BC048679	7	6.7	2.8
Mgll	6	6.6	4.7
Dlk1	12	6.3	9.3
Itga11	9	6.3	47.7
9930105H17Rik	ND	6.3	3.3
B230397F11Rik	ND	6.2	(1)
LOC329416	ND	6.1	2.7

***Table 5.1*. Transcripts significantly de-repressed in *Kaiso*-null fibroblasts**. Shown is the chromosomal location of each transcript and the fold expression change in *Kaiso*-null and MKO fibroblasts relative to wild-type. A bracketed number one (1) indicates expression was not detected in either wild-type or MKO cells and therefore there was no expression change.

To investigate whether any biological ontologies were significantly enriched among genes upregulated in *Kaiso*$^{-/-}$ and MKO fibroblasts, I used the DAVID v6 programme, employing the same parameters as previous analyses (*Section 3.5*). I found that genes upregulated in the absence of Kaiso were enriched for the categories biological process (p = 0.0012) and developmental process (p = 0.02) at the hierarchical level biological process 1 (BP1) (*Fig 5.2c*). Moreover, these genes

were also enriched for brain specific expression (p = 0.046), consistent with strong Kaiso expression in brain tissues (Della Ragione et al., 2006). In contrast, genes upregulated in MKO fibroblasts were strongly enriched for immune system processes ($p = 8^{-10}$) and also developmental process (p = 0.000052). Consistent with this strong enrichment of genes associated with the immune system, MKO upregulated genes were preferentially expressed in activated spleen (p = 0.00083), bone marrow (p = 0.0022), spleen (p = 0.0026) and also interestingly, mammary tumours (p = 0.015). The strong enrichments for immune associated ontology's and tissues suggests that MKO cells may be inherently stressed and/or responding immunologically. However, MKO cells appear morphologically normal and proliferate comparably to wild-type parental fibroblasts. Thus, it is also possible that MeCP2, MBD2 and Kaiso have specific or degenerate roles in regulating immune and stress-response genes, although this would not necessarily be a methylation-dependent function.

Notably, gene ontology analysis suggested neither *Kaiso*-null nor MKO fibroblasts were enriched in expression of germline-specific transcripts (*Fig 5.2c*), supporting the earlier conclusion that MBPs are not the preferential mechanism that mediates repression of methylated germline genes (*Fig 5.2a*). However intriguingly, the genes upregulated in Kaiso-deficient cells exhibited a striking level of tissue-specificity, but not for any particular tissue. For example upregulated *Bcl11b* is normally specifically expressed in the thymus, *Zic1* only in the cerebellum, *Megf6* exclusively in osteoblasts, *Lhx9* in the hippocampus and *Klhl10* specifically in testis (*Table 5.1*). This general tissue-specificity of upregulated genes may hint at a role for Kaiso in fine tuning tissue-specific expression of putative target genes, particularly in the brain, which is significantly enriched (p = 0.046). This function could potentially be methylation-dependent or regulated by alternative mechanisms such as tissue-specific canonical Wnt signalling, or both (Ruzov et al., 2009b).

Because DNA methylation is a crucial component of the regulation of genomic imprinting, it has been proposed that MBP deficiency might lead to dysregulation of imprinted gene expression (Bird, 2002). To explore the possibility that Kaiso

mediates methylation-dependent regulation of imprinted loci, I determined whether imprinted genes were upregulated greater than 1.5-fold in the absence of Kaiso. I imposed this 1.5-fold threshold as it would be predicted that one imprinted allele would already be expressed (depending on cell-type) and therefore the theoretical maximum upregulation would be 2-fold i.e. de-repression of the second allele. This analysis found 6 imprinted genes were de-repressed >1.5-fold in *Kaiso*-null fibroblasts (*Table 5.2*). However, using the premise that there are 131 imprinted loci (MRC Harwell Mousebook), it would be expected that 9 imprinted genes would be found in this dataset by chance, indicating Kaiso is not a general regulator of imprinted gene expression. Nonetheless, this does not preclude the possibility that Kaiso has locus-specific functions at the imprinted genes identified here. It would be interesting to investigate this possibility in future studies.

Imprinted gene	Chromosome	Expression allele	Fold-change in Kaiso-/-
Dlk1	12	P	9.3
Snurf	7	P	3.2
Magel2	7	P	2.2
Mest	6	P	2
Mkrn3	7	P	1.8
Peg13	15	P	1.8

***Table 5.2*. Imprinted genes dysregulated in *Kaiso*-null fibroblasts**. Shown is the chromosomal location of each imprinted locus and the actively transcribed allele where P: paternal allele and M: maternal allele. Fold changes represent expression values in *Kaiso*-null cells relative to wild-type.

It has been reported that Kaiso is a bimodal transcriptional repressor that can bind both methylated target genes and CTGCNA motifs (Prokhortchouk et al., 2001; Daniel et al., 2002; Kim et al., 2004b). To identify the potential methylation-dependent targets of Kaiso, I cross-referenced genes de-repressed in Kaiso-null cells and genes de-repressed in hypomethylated DP cells. It would be predicted that genes critically reliant on the methyl-CpG binding function of Kaiso for repression would be de-repressed in both datasets. However, genes de-repressed exclusively in Kaiso-null cells and not DP cells would be predicted to be regulated via methylation-independent Kaiso functions or as a consequence of indirect effects. Because the Ref-6 BeadChip used for analysis of Kaiso-null fibroblasts contained an additional

12,668 transcripts to the Ref-8 BeadChip used for DP cell analysis, I excluded these additional transcripts from this analysis. This exclusion left 21 genes (of 31) common between both datasets and upregulated >6-fold in Kaiso-null cells. Of these, I found that only one gene, *Tex19,* was also de-repressed >6-fold in DP cells. This suggests that Kaiso may only have one physiological methylation-dependent target gene. However, this conclusion may reflect the stringent thresholds used here and not a unique biological Kaiso target. For example, when the threshold value is altered to include genes expressed >4-fold in *Kaiso*-null and DP cells, 6 genes are present in both datasets. Indeed, when transcripts upregulated >2 fold in Kaiso-null cells are cross-referenced with transcripts upregulated >6-fold in DP cells there are 14 common genes and when both datasets are set to a >2-fold threshold, 50 genes are de-repressed in both (*Fig 5.2e*). This suggests that there could potentially be multiple methylation-dependent Kaiso target genes. Indeed, the relatively strong activation of *Tex19* in both Kaiso-deficient and DP cells compared to other de-repressed genes (and hence the only gene activated >6-fold in both datasets) could be accounted for by its inherently strong promoter (*Section 4.4*) and not because it is the preferential Kaiso target. However, this study indicates that *Tex19* probably does represent a relatively rare and specific methylation-dependent Kaiso target – the proverbial 'needle in a haystack'.

In summary, this global expression analysis has demonstrated that only a modest number of genes are significantly de-repressed in Kaiso-deficient fibroblasts. The lack of general gene activation is consistent with genetic studies that suggest MBPs do not act as global regulators of gene expression. However, the data is also consistent with Kaiso potentially contributing to gene silencing at a number of loci and importantly, at a few limited genes (*Table 5.1*) being a primary mechanism through which methylation dependent silencing is mediated. *Tex19* falls into this latter category as my expression array suggests it is strongly de-repressed in Kaiso-null, MKO and DP cells. It is likely that the genes de-repressed in *Kaiso*-null and *Dnmt1*-null cells, such as *Tex19,* represent a small minority of loci that rely exclusively on DNA methylation and Kaiso for silencing. This data thus supports the

argument that *Tex19* is a specific and rare target for methylation-dependent silencing by Kaiso.

5.4 Exogenous Kaiso can rescue *Tex19* de-repression and is essential for silencing during ES cell differentiation

The present study has indicated *Tex19* is de-repressed but still methylated in the absence of Kaiso indicating Kaiso couples methylation to transcriptional silencing at this locus. Moreover, global expression data suggests *Tex19* is a specific Kaiso-target on a genome-wide scale. To confirm that loss of Kaiso is the causal factor for *Tex19* de-repression, I examined whether expression of exogenous Kaiso could rescue *Tex19* silencing. I transiently transfected GFP-tagged *Xenopus* Kaiso (xKaiso) into *Kaiso*$^{-/-}$ fibroblasts for 24hrs and flow sorted cells according to GFP expression. qRT-PCR analysis of GFP positive (xKaiso rescued) and GFP negative (control) cells demonstrated that transient expression of xKaiso can re-impose *Tex19* repression by ~5-fold (*Fig 5.3a*). Thus, *Xenopus* Kaiso, which retains methyl-binding activity, can partially rescue the de-repression of *Tex19* as a result of mouse Kaiso deficiency. This strongly supports the argument that loss of Kaiso *per se* is the cause of *Tex19* activation in *Kaiso*-null fibroblasts

To determine the precise importance of Kaiso for initiating developmental silencing of *Tex19,* I collaborated with the *Prokhortchouk* laboratory. Here, I was able to analyse *Tex19* expression in *Kaiso*$^{-/-}$ ES cells during differentiation into embryoid bodies (EB). qRT-PCR demonstrated that wild-type parental ES cells fully silenced *Tex19* expression during EB formation comparably to wild type E14 and J1 ES cells used previously (*Section 4.7* & *4.8*). Strikingly, Kaiso-deficient ES cells were unable to impose *Tex19* silencing during EB formation. Following 9 days EB differentiation, *Tex19* was still expressed at ~25% of the levels at day 0 in Kaiso-null EBs compared to 0.002% in wild-type EBs (*Fig 5.3b*). Both ES cell lines were able to silence *Oct3/4* expression comparably during EB formation indicating they both underwent equivalent programmes of differentiation (*Fig 5.3c*) and consistent with previous reports that *Kaiso*-null ES cells differentiate efficiently (Prokhortchouk et

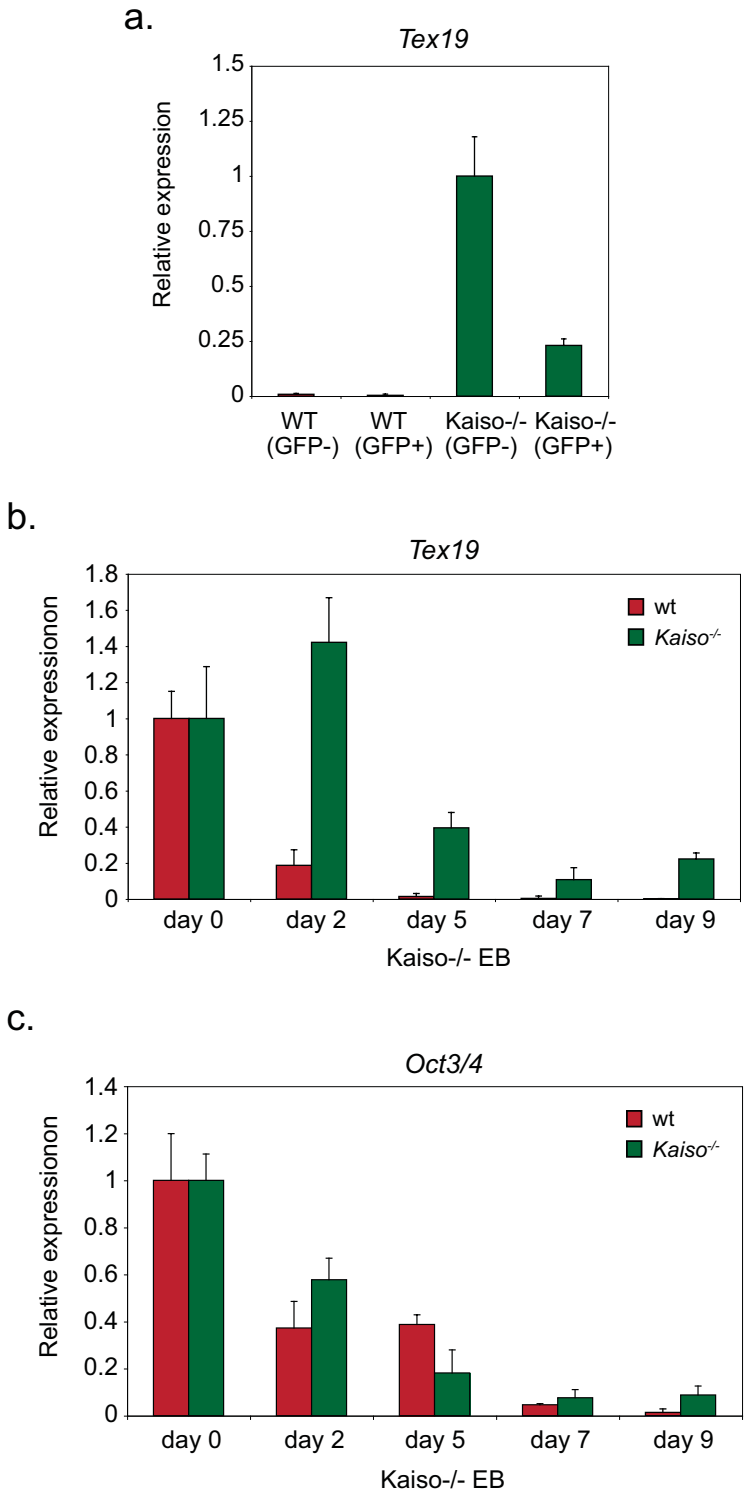


Fig 5.3. **Kaiso is required to silence *Tex19* in differentiating ES cells and partially rescues repression in *Kaiso*-null fibroblasts.** a.) qRT-PCR expression analysis of *Tex19* in wild-type and *Kaiso*-mutant fibroblasts transiently transfected and FACS sorted for GFP expressing xKaiso (GFP+) or untransfected (GFP-). qRT-PCR of b.) *Tex19* and c.) *Oct3/4* expression in wild-type and *Kaiso*-null ES cells during embryoid body formation. b.) *Tex19* is strongly silenced in wt EBs but fails to impose silencing in the absence of Kaiso. c.) *Oct3/4* is silenced comparably between differentiated wt and *Kaiso*-null ES cells. Shown is relative expression to undifferentiated ES cells (day 0), set to 1. Expression is normalised to *Gapdh.*

al., 2006). These data strongly support a critical role for Kaiso in mediating initiation and maintenance of gene silencing at *Tex19.* Considered with the data presented in previous chapters (*Fig 4.8 & 4.9*), this work suggests that Kaiso is the critical factor that couples promoter methylation at *Tex19* to downstream silencing.

5.5 Kaiso is localised to the methylated *Tex19* promoter

Analysis of the *Tex19* promoter showed that it is enriched with the symmetrically methylated CGCG tetranucleotides that Kaiso preferentially binds to in a methylation-dependent manner, and that these are located adjacent to the TSS (*Fig 4.1a & 4.2*) (Prokhortchouk et al., 2001). In contrast, Kaisio binding sequences (KBS) were not detected within 20kb of the *Tex19* TSS (*see Section 1.4.6.1*) (Daniel et al., 2002). This suggests that Kaiso could strongly bind *Tex19* at methylated CGCG motifs indicating a potential mechanism for targeting Kaiso specifically to this locus. To confirm that Kaiso is localised to the *Tex19* promoter in methylated somatic cells, I performed ChIP of endogenous Kaiso protein with a monoclonal α-Kaiso antibody (Abcam ab12723) in wild-type tailtip fibroblasts. Optimisation protocols demonstrated that, in my hands, this antibody had highly limited avidity for mouse Kaiso and/or was not ChIP grade, as judged by enrichments for putative positive control genes. Despite this, successive experiments demonstrated that Kaiso was consistently, but modestly, enriched at the methylated *Tex19* promoter in wild-type fibroblasts (*Fig 5.4a*). No enrichment of Kaiso at *Tex19* was detected in *Kaiso*$^{-/-}$ fibroblasts suggesting enrichment in wild type cells was a specific effect.

To confirm that Kaiso specifically targeted only methylated *Tex19,* I explored other approaches. I created a fusion protein between the zinc-finger domain of *Xenopus* Kaiso, which is responsible for methyl-binding, and the viral VP16 activation domain, which strongly activates transcription of promoters to which it is localised (xZF-VP16). I reasoned that the xZF-VP16 fusion protein would specifically activate a methylated *Tex19* reporter relative to VP16, as only xZF-VP16 would putatively bind methylated *Tex19.* In contrast, the ratio of activation between VP16 and xZF-

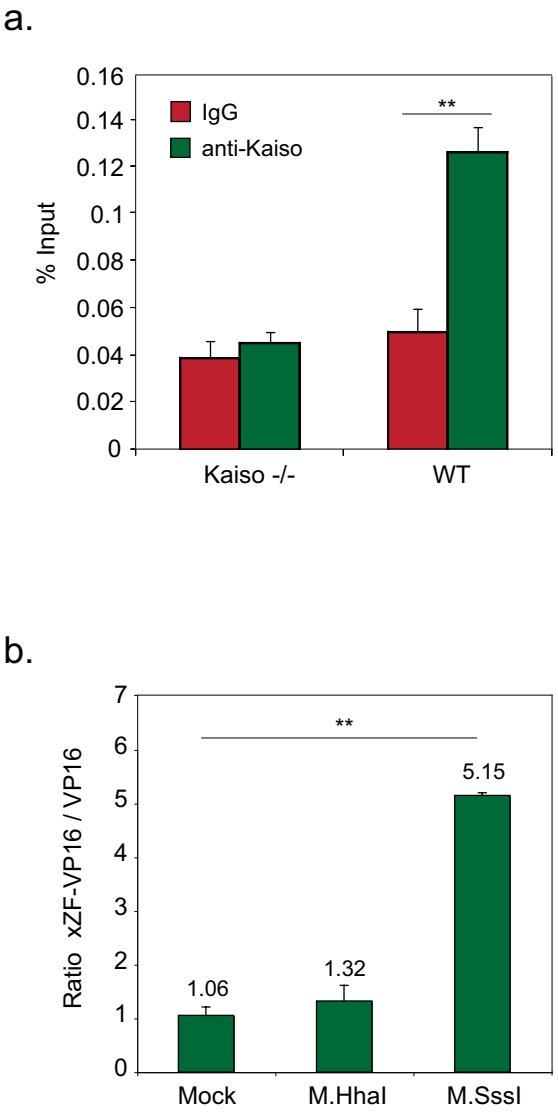


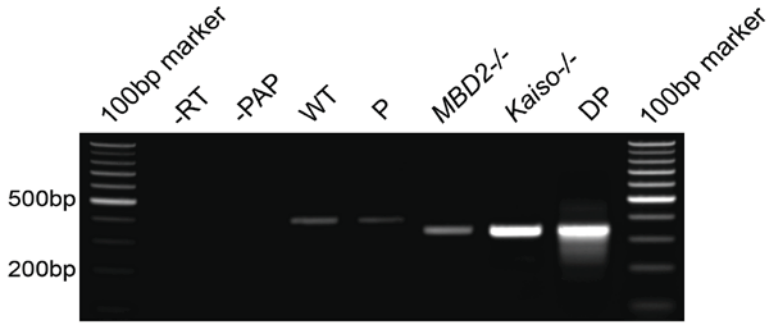
Fig. 5.4. **Kaiso is recruited to methylated *Tex19* promoter.** a.) ChIP of endogenous Kaiso from control *Kaiso*-null and wild-type fibroblasts. Enrichments were determined by qRT-PCR of the *Tex19* promoter region. Error bars are S.E.M b.) Kaiso zinc-finger domain only binds methylated *Tex19* reporter. Unmethylated or methylated *Tex19* luciferase reporter was transiently transfected into 293T cells with either xZF-VP16 or VP16. xZF-VP16 specifically activates methylated (*Sss.I*) *Tex19,* shown as a ratio to VP16 alone. xZF-VP16 does not significantly activate mock methylated or partially methylated (*Hha.I*) *Tex19* relative to VP16. Shown are normalised luciferase values + SEM. ** = p <0.01.

VP16 at unmethylated *Tex19* reporter would be predicted to be ~1 as neither would preferentially bind. Transient transfection of these constructs into 293T cells demonstrated that xZF-VP16 strongly activated fully methylated (Sss.I) reporter at a ratio of 5.15-fold relative to VP16 (*Fig 5.4b*). However as predicted, the ratio of activation at unmethylated *Tex19* was 1.06, suggesting neither VP16 nor xZF-VP16 preferentially localised to the unmethylated reporter. Consistent with Kaiso having weak binding activity for methylated CpGs in a non-CGCG context, partially methylated reporters (*Hha.I*) only exhibited a modest ratio shift towards xZF-VP16 preference (ratio = 1.32) (*Fig 5.4b*). This data strongly suggests that Kaiso binding activity can strongly localise to *Tex19,* but only when methylated. This supports the argument that Kaiso is responsible for mediating the methylation-dependent silencing of *Tex19* expression.

5.6 *Tex19* is transcribed from the canonical promoter in Kaiso-null cells

Potentially, *Tex19* expression in *Kaiso*-null and MKO fibroblasts could initiate from an alternative promoter through a Kaiso-independent mechanism, which would thus account for de-repression in MBP mutant cells. To evaluate this possibility, I used a 5' rapid amplification of cDNA ends (5'-RACE) assay to map the transcriptional start sites (TSS) of *Tex19* in multiple cell types. The primers which I designed were predicted to generate a 314bp product if *Tex19* transcription originated from the canonical annotated TSS (RefSeq - http://www.ncbi.nlm.nih.gov/refseq/ & DataBaseTSS - http://dbtss. hgc.jp/). I found that *MBD2*$^{-/-}$, *Kaiso*$^{-/-}$, and DP cells all generated a band of ~314bp that, as expected, showed proportional intensity to the levels of expression in the cells i.e. highest in DP and lowest in *MBD2*$^{-/-}$cells (*Fig 5.5a*). This suggests that DP, *MBD2*-null and *Kaiso*-null cells all transcribe *Tex19* from the canonical TSS and therefore have lost silencing of *Tex19* rather than utilised alternative transcription start sites. Interestingly, overcycling the 5'RACE reactions generated a weak product from both P cells and wild-type fibroblasts, that appeared larger than expected (*Fig 5.5a*). This may reflect weak basal expression that initiates from an alternative promoter in non-expressing cell types.

a.



b.

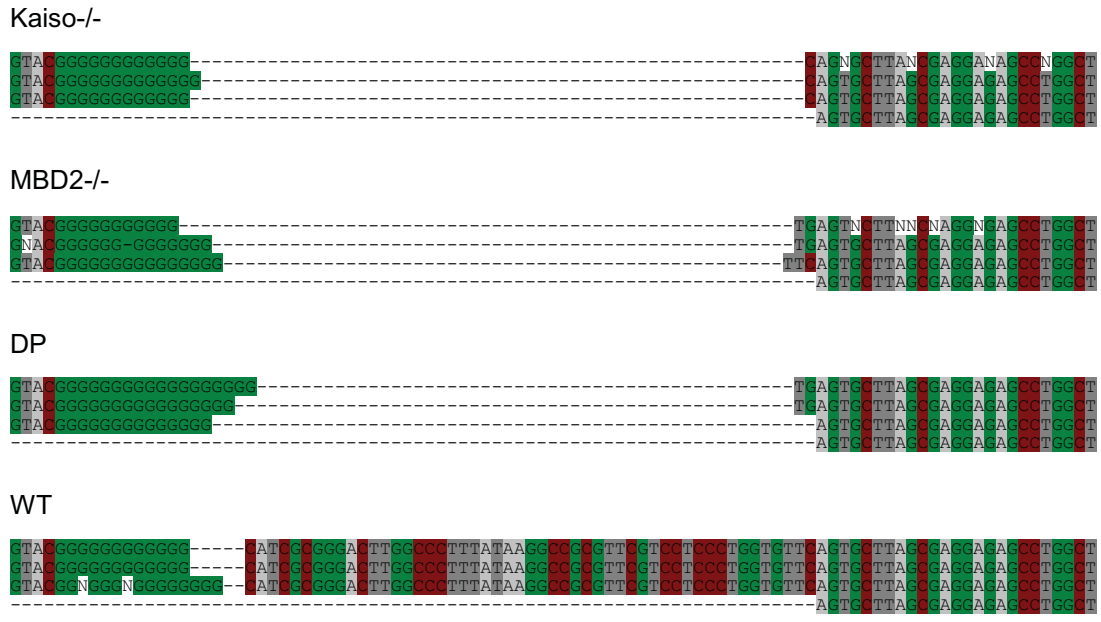


***Fig 5.5*. Tex19 is transcribed from the annotated promoter in mutant cells.** a.) Products generated from 5' rapid amplification of cDNA ends assay (5'RACE). A band of ~314 bp indicates *Tex19* transcription initiates from the annotated canonical promoter. b.) Sequencing of the products generated in 5.5a. Three independent sequenced clones are shown with the expected 'canonical' AGT start site shown below. Sequencing confirmed that *Tex19* is transcribed from the expected 'AGT' start site in DP, *Kaiso*-null, and *MBD2*-null cells. In contrast, weak transcription initiates ~51bp upstream of this in non-expressing wild-type cells. The poly-G residues (left) are a consequence of the cloning strategy.

To confirm that expression in mutant cells occurs from the canonical TSS and investigate the weak transcript in P and wild-type cells I sequenced the 5'-RACE products. This analysis confirmed that in the absence of DNA methylation (DP) or MBPs (Kaiso and MBD2), *Tex19* is de-repressed and that transcription initiates from the annotated 'AGT' TSS (*Fig 5.5b*). This supports a conclusion whereby loss of promoter methylation or Kaiso deficiency is sufficient to allow active transcription from the canonical *Tex19* promoter and thus suggests that these components critically promote gene silencing and regulate expression patterns. This data rules out the possibility that de-repression of *Tex19* in *Kaiso*-null fibroblasts occurs due to alternative promoter usage. Intriguingly, the weak transcript detectable (with overcycling) in non-expressing P cells and wild-type fibroblasts initiated from ~51bp upstream of the canonical *Tex19* 'AGT' TSS (*Fig 5.5b*). This suggests that any background basal expression from this locus is unable to initiate from the native TSS, presumably due to inhibition by promoter methylation and Kaiso, but is capable of weak transcriptional initiation from an upstream cryptic TSS. Notably this cryptic transcript is present at >4000-fold lower levels than transcripts generated from the canonical unmethylated TSS and is therefore likely not biologically relevant (*Fig 3.10b*). Importantly, this analysis confirms the key role of Kaiso in interpreting *Tex19* methylation and coupling transcriptional silencing to the canonical promoter in a methyl-dependent manner.

5.7 An antisense transcript is transcribed through the *Tex19* coding region and promoter

An intriguing aspect of this work is that *Tex19* (and several other germline candidate genes) are targeted for promoter methylation in somatic cells *in vivo,* despite being CGI promoters (HCPs) that would normally avoid *de novo* methylation. The mechanism that identifies *Tex19* as a target for DNA methylation yet excludes the majority of other CGIs is of great interest for understanding the regulation of this locus and indeed for determining how CGIs are excluded from *de novo* CpG methylation in general. Previous studies have noted that antisense transcripts can affect transcription in *cis* both negatively (Yu et al., 2008) and positively (Scheele et

al., 2007) and can target DNA methylation to promoters in *cis* (Chotalia et a., 2009). To investigate whether antisense transcripts could have a role in regulating *Tex19* through guiding DNA methylation to the promoter, I initially interrogated the NCBI EST database. This analysis found that an antisense transcript complimentary to the upstream proximal promoter of *Tex19* was expressed specifically in testis. To validate expression of this putative antisense transcript in testis and define its size and origin I designed antisense qRT-PCR primers both upstream and downstream of the *Tex19* coding region (*Fig 5.6a*). I also designed antisense primers to a CpG island 8kb downstream of *Tex19.* qRT-PCR analysis of testis detected strong expression of an antisense transcript 200bp downstream of the 3' end of *Tex19* (3') and also at the 5' *Tex19* promoter (Prox). This expression declined ~4-fold 600bp upstream of the promoter (Dis) indicating the antisense transcript, at least partially, terminates transcription after reading through the *Tex19* promoter. No expression was detected at the distal region 8kb downstream of *Tex19* (CpG), adjacent to the CpG island (*Fig 5.6b*). This data suggests that expression of an antisense transcript (*asTex19*) initiates between the 3'end of *Tex19* and 8kb downstream of this and that it is constitutively transcribed through the entire *Tex19* coding and promoter region in testis.

It has been postulated that antisense transcripts could function to guide DNA methylation to specific sites (Bernstein & Allis, 2005). Moreover, work at maternally imprinted loci has demonstrated antisense transcription establishes an open chromatin domains that allow DNA methylation machinery access to *de novo* methylate the region (Chotalia et al., 2009). However, in the present study expression of *asTex19* seems to occur in testis, precisely where *Tex19* is protected from DNA methylation. Thus, if *asTex19* were to have a functional epigenetic role, it would not be in targeting *Tex19* promoter methylation but a novel role in protecting it from methylation. To investigate this possibility I examined *asTex19* expression in wild-type, *Kaiso*-null, and DP cells. Because these cell-types each have a unique combination of *Tex19* methylation and expression status, I would be able to correlate the expression of *asTex19* to downstream effects i.e. *Tex19* methylation and expression. This analysis demonstrated that expression of *asTex19* could not be detected in wild-type fibroblasts but was readily detected at all loci (Dis, Prox, CpG)

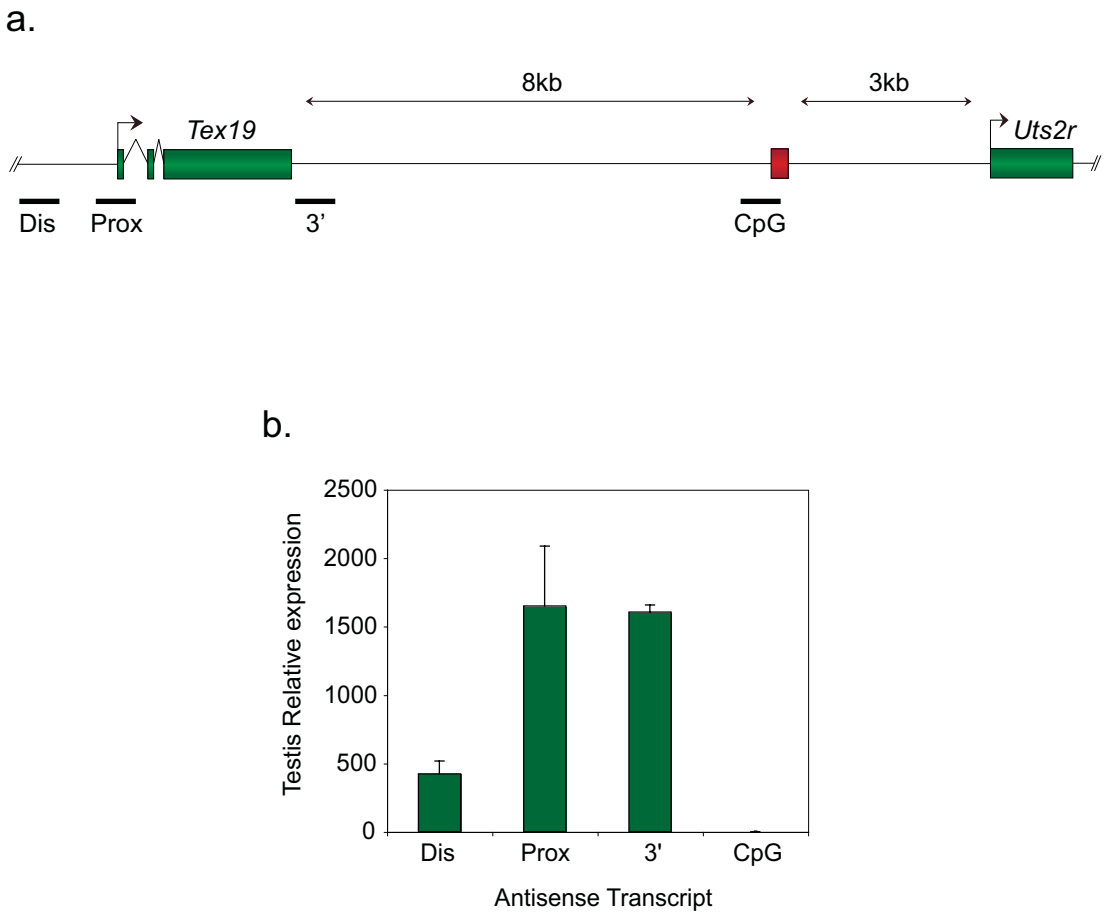


Fig. 5.6. **Identification of antisense *Tex19* transcripts in testis.** a.) Schematic of regions amplified to identify the antisense transcript. The upstream and downstream antisense regions tested for expression are shown as black bars. The *Tex19* exon structure is shown and the TSS marked by arrow. A downstream CpG island (CGI) is marked in red. b.) qRT-PCR expression analysis of the four antisense regions in testis. Expression is relative to CpG, which is set to 1. All expression is normalised to *Gapdh*.

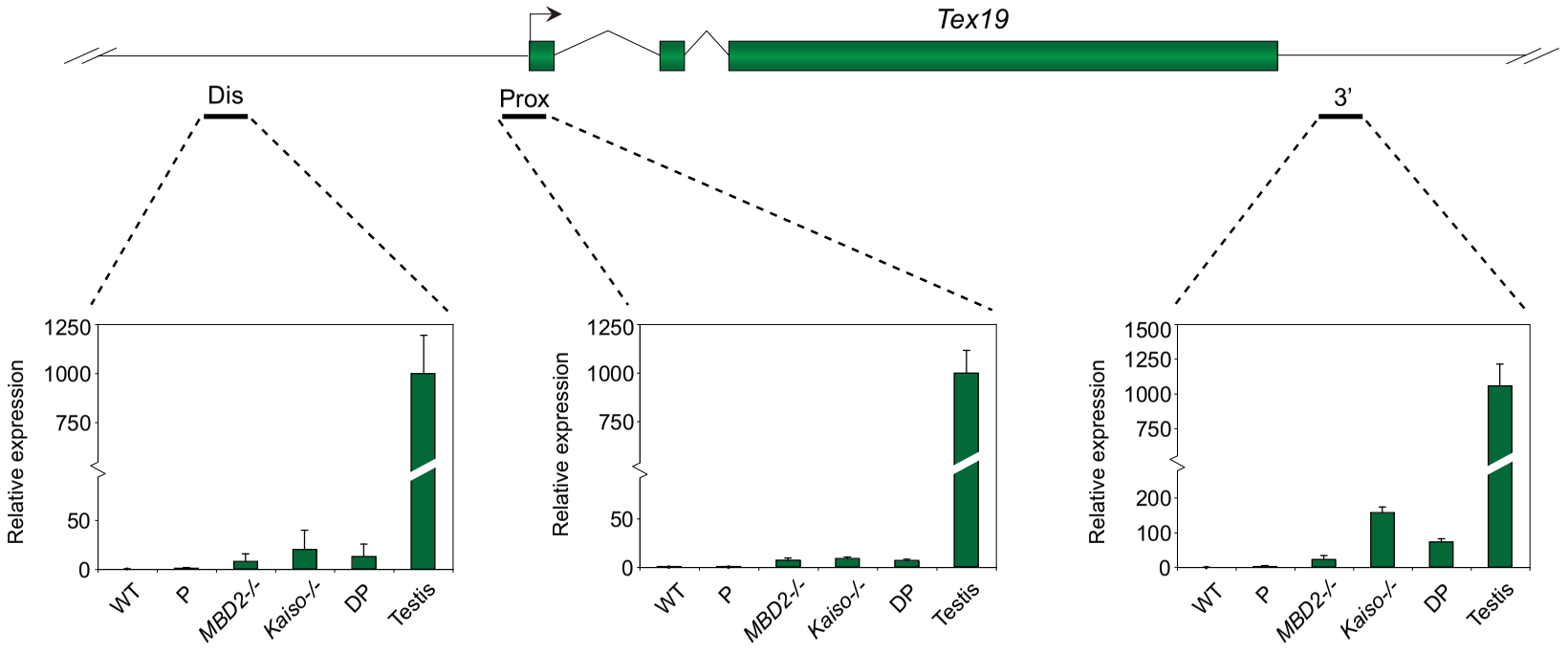


***Fig. 5.7*. Antisense transcripts are expressed in *Tex19* expressing cell types.** qRT-PCR expression analysis of *asTex19* from three genomic regions + S.E.M. *asTex19* is only expressed in cell-types that express *Tex19* and is not correllated with the promoter methylation status of *Tex19*. Expression is relative to WT, which is set to 1 and is normalised to *Gapdh.*

in *Kaiso*-null and *DP* cells, although not to the same extent as testis (*Fig 5.7*). Importantly, as both *Kaiso*-null and DP cells express *Tex19* but only DP cells are hypomethylated this suggests that *asTex19* transcription is correlated with *Tex19* expression and not *Tex19* methylation status. This conclusion is supported by the observation that *MBD2*-null cells express *asTex19* but P cells do not (*Fig 5.7*). Taken together, this data suggests that antisense transcription through *Tex19* is a consequence of expression of *Tex19* itself and not a regulatory mechanism *per se*.

One conceptual possibility is that *Tex19* transcription generates a permissive or open downstream chromatin domain that allows activation of a cryptic antisense promoter and thus transcription of *asTex19.* This is supported by the observation that *asTex19* is expressed at comparable levels in *MBD2*-null, *Kaiso*-null and DP cells despite the disparate range of expression of *Tex19* between these cells. Here the argument suggests that transcription of *Tex19*, at any level, would cause full activation of *asTex19.* Indeed, in turn *asTex19* could potentially have a role in targeting gene body methylation to the *Tex19* coding region, which has been associated with gene activity and is thought to prevent cryptic transcriptional initiation (Eckhardt et al., 2006; Suzuki & Bird, 2008). Interestingly, a cryptic transcriptional start site was observed in wild-type and P cells, both of which do not express *asTex19,* but this cryptic transcript was not present in cell types that did express *asTex19.* However, no evidence for any regulatory function for *asTex19* has been generated here and future work will be required delineate the role of this antisense transcript*,* if any. In summary, the data generated here is consistent with *asTex19* expression being a downstream consequence of *Tex19* activation and therefore not a primary regulator of *Tex19* methylation patterns.

5.8 The *Tex19* promoter is characterised by retroelements.

My studies suggested that an antisense mediated mechanism was not responsible for the observed *Tex19* methylation patterns. To explore alternative mechanisms that potentially target methylation to the *Tex19* CGI promoter, I investigated the promoter sequence structure using *RepeatMasker* (http://www.repeatmasker.org/).

Retroelement derived sequences are present at ~13% of mouse proximal promoters and have been shown to have dramatic epigenetic effects on adjacent genes (Tomilin, 2008). For example, LINE1 elements can induce extensive DNA methylation and silencing of local genes whereas in some contexts SINEs have been shown to block this effect by functioning as insulators against the spreading of DNA methylation and chromatin compaction (Noma et al., 2006). Analysis of *Tex19* demonstrated that a significant proportion (~49%) of the extended promoter (-2500bp – TSS) was composed of retroelement derived sequences (*Fig 5.8*). Across this region 26.3% were short interspersed nucleotide elements (SINEs), 9.1% long interspersed nucleotide elements (LINES) and 13.4% were long terminal repeat derived elements (LTRs). No DNA elements were detected.

Following their insertion into DNA, LINE and LTR elements are targeted for heritable methylation to ensure their transcriptional silencing. The 'genome defence' hypothesis argues that DNA methylation evolved primarily for this purpose (Yoder et al., 1997). However, it is also argued that insertion of retroelements can drive evolution of the expression pattern of adjacent genes. One possibility is that the *Tex19* promoter is targeted for DNA methylation, despite being a CGI, because of a spreading effect from local retroelement methylation. LINEs and LTRs are highly depleted at mouse promoters (4.34% at promoters' vs 27.92% in the genome) and thus their strong enrichment at *Tex19* may reflect a distinctive role, such as targeting methylation to the CGI promoter in somatic cells (Tomilin, 2008).

Further analysis of *Tex19* demonstrated a SINE element was also present ~2000bp downstream of the TSS and thus the *Tex19* promoter is flanked by SINE elements (*Fig 5.8*). While SINES are also often *de novo* methylated, this opens up the intriguing possibility that the flanking SINEs may act as epigenetic insulators against chromatin spreading analogous to the mechanism reported in *S. pombe* and at the mouse growth hormone locus (Noma et al., 2006; Lunyak et al., 2007). Here, CpG-rich sequences that are flanked by SINE elements have been shown to be protected from repressive histone modifications. Taken together these observations hint at a potential but highly speculative mechanism for guiding CpG methylation to *Tex19*

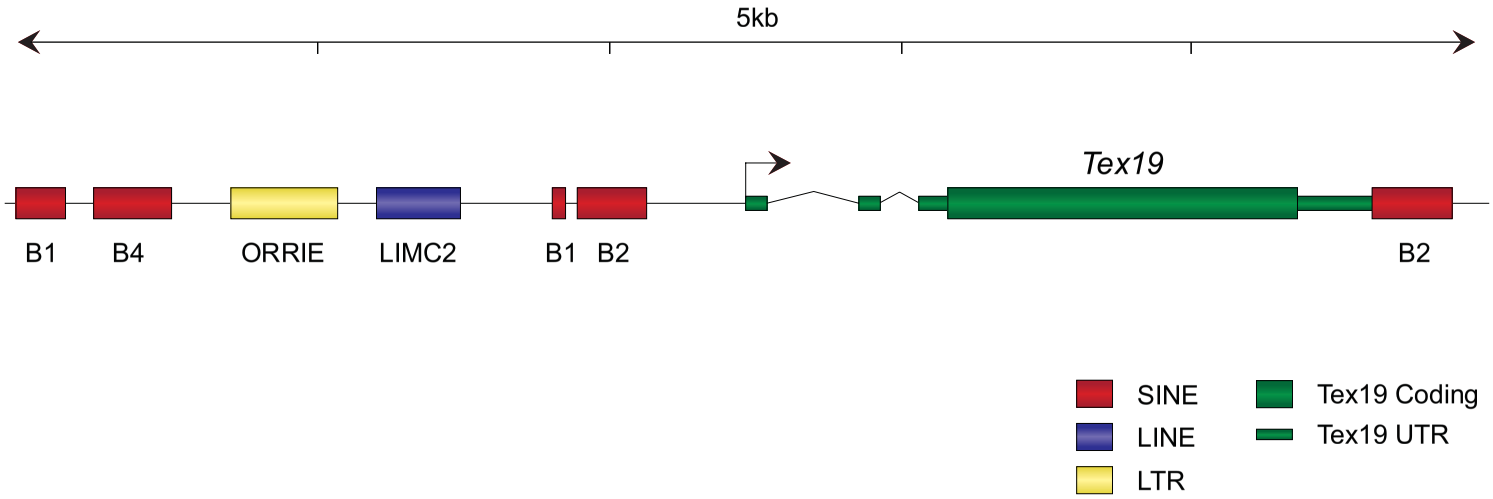


Fig 5.8. **Schematic of the retroelements present in the proximal *Tex19* promoter.** Forty nine percent of the upstream promoter (-2500bp-TSS) is composed of retroelements which may have a role in targeting *de novo* methylation to the locus. Note the flanking B2 SINE elements either side of the *Tex19* transcript which could potentially have a role in insulating the region from local chromatin structure, including methylation in the germline. The subclass of retroelement is shown under each element.

(targeting by LINEs and LTRs) and also suggest why histone modifications are depleted at the locus (insulation by flanking SINEs). However, further work to dissect the precise mechanism that developmentally targets methylation to *Tex19* is crucial.

5.9 Discussion

Despite considerable research, the *in vivo* mechanism of methylation-dependent gene silencing remains unclear. A likely scenario is that at many loci, both epigenetic modifications and DNA methylation contribute to transcriptional silencing. At these loci the CpG methylation portion of repression may be mediated, at least in part, through methyl-binding proteins. However, importantly other epigenetic marks and/or trans-acting factors would also be predicted to maintain silencing independently of CpG methylation. That is to say, DNA methylation is generally a secondary epigenetic system that reinforces epigenetic states but does not direct them. Therefore, abrogation of a methyl-binding protein would not be predicted to significantly affect transcription as loss of DNA methylation *per se* would not directly affect transcription levels at most loci (although there may be indirect effects). Unless DNA methylation were the primary mechanism of silencing at a gene locus, it would be predicted that methyl-binding proteins would be dispensable for gene repression. Indeed, genetic studies have demonstrated that at the phenotypic and molecular expression level, MBPs are dispensable. Conceptually this is not surprising given that currently, there are very few, if any, examples of genes that are regulated primarily and exclusively by promoter CpG methylation. As noted earlier, most genes are thought to be regulated by multiple layers of epigenetic and trans-acting systems. Therefore, to identify genes regulated by MBPs it is also necessary to identify genes regulated exclusively, or at least primarily, by DNA methylation.

In this thesis I identified a set of germline-specific genes that were candidates for direct regulation by promoter CpG methylation. Further characterisation of two candidates demonstrated *Tex19* was exclusively reliant on DNA methylation for silencing. Because of this crucial requirement for promoter methylation to maintain

repression, I investigated whether MBPs were responsible for imposing methyl-dependent silencing at *Tex19*. This analysis demonstrated that in the absence of Kaiso, *Tex19* was highly de-repressed, yet retained full promoter methylation. This strongly suggests that Kaiso is critically responsible for coupling promoter CpG methylation at *Tex19* to transcriptional silencing. However, whilst *Tex19* lost silencing in *Kaiso*-null cells, it remained partially repressed compared to hypomethylated DP cells. This is consistent with Kaiso being required to fully silence *Tex19* expression but also with CpG methylation *per se* significantly repressing transcription, either directly or through an alternative mechanism such as protein exclusion or recruitment of other MBPs. An interesting possibility here is that the other two Kaiso-like proteins, ZBTB4 and ZBTB38 may partially compensate for loss of Kaiso (Filion et al., 2006). siRNA knockdown studies of these proteins in *Kaiso*-null fibroblasts will be important to determine if *Tex19* can be de-repressed further. Importantly, analysis of Kaiso-null ES cells demonstrated silencing could not be imposed upon *Tex19* during differentiation, indicating that Kaiso is crucially required for initiation, in addition to maintenance, of *Tex19* repression. Moreover exogenous expression of xKaiso could partially re-impose *Tex19* repression in Kaiso-deficient fibroblasts. Further studies confirmed that endogenous Kaiso is recruited to the *Tex19* promoter, which contains several Kaiso methyl binding sites. Indeed, Kaiso association with *Tex19* depended on the promoter being methylated. Taken together these data provide compelling evidence that *Tex19* is a methylation-dependent Kaiso target. This thesis represents a novel report of gene that relies on both promoter methylation for gene silencing and Kaiso to mediate the downstream silencing effects of CpG methylation.

Interestingly, global expression analysis suggested that *Tex19* was one of a highly limited number of methylation-dependent genes de-repressed in Kaiso-null cells and therefore a highly specific target. This is consistent with the relatively small number of genes that rely primarily on DNA methylation and also previous Kaiso studies (S*ection 3.9*) (Prokhortchouk et al., 2006). Ongoing analyses will confirm whether the remaining putative Kaiso targets identified here, are direct or indirect/false-positive hits.

Chapter 6

Discussion

6.1 Promoter CpG methylation regulates germline-specific gene expression

It has been demonstrated by multiple studies that lineage-specific genes are co-ordinately regulated by epigenetic systems and trans-acting factors which impose heritable transcriptional memories (Chen et al., 1998; Feldman et al., 2006; McCarrey et al., 2006; Ng et al., 2008; Hemburger et al., 2009; Jaenisch and Bird, 2003). In contrast, it has been proposed that a subset of germline-specific genes are regulated primarily by DNA methylation *per se* (Oda et al., 2006; Maatouk et al., 2006; Weber et al., 2007). A fundamental drawback of this proposal is that it is only supported by correlative or indirect evidence and thus remains to be proven. In the present study, I provide novel cause and effect data comprehensively demonstrating that at least one germline-specific gene, *Tex19,* is regulated primarily and exclusively by DNA methylation. Additionally I have identified a novel set of germline-specific candidate genes that putatively rely on promoter DNA methylation analogously to *Tex19.* This work provides the first causal evidence that CpG methylation directly regulates gene expression exclusive of tissue-specific trans-acting factors and alternative epigenetic systems. This affords novel proof of principle that promoter DNA methylation could be a general mechanism for regulating a subset of germline-specific genes.

I investigated the role of DNA methylation by examining the transcriptional effect of cellular demethylation through three different experimental approaches – (i) Dnmt1 inactivation, (ii) 5-aza dC treatment and (iii) recovery following 5-aza dC treatment (*Section 3.5, 3.6 & 3.7*). I found germ cell-associated genes were highly enriched among upregulated transcripts in each approach, and by cross-referencing the experiments, I identified 14 genes (9 testis-specific) as candidates for regulation by

DNA methylation (*Section 3.8*). Although most CpG islands (CGI) are unmethylated regardless of tissue or expression status (Rollins et al., 2006; Weber et al., 2007), I found each CGI candidate gene tested was highly methylated in somatic cells but demethylated in expressing germ cells (*Section 4.2*). Moreover, candidate promoters could drive strong expression in somatic cells, suggesting non-expressing cells carry activating transcription factors and therefore that an epigenetic system (putatively promoter methylation) must mediate silencing of endogenous candidate loci (*Section 4.4*). Indeed, *in vitro* methylation of candidate promoters caused strong repression of reporter transcription. To test whether DNA methylation was the upstream epigenetic system repressing the candidate gene *Tex19,* I investigated promoter methylation dynamics during ES cell differentiation. Importantly, I found that DNA methylation accumulated at the *Tex19* promoter coincident with or prior to gene silencing i.e. was functionally upstream of silencing (*Section 4.7*). To confirm CpG methylation was the cause of repression (rather than just correlated with it) I demonstrated that inactivation of the *de novo* methyl-transferase Dnmt3b suppressed *Tex19* silencing in differentiated ES cells. These data suggest that *de novo* DNA methylation is sufficient and necessary for *Tex19* silencing (*Section 4.8*). Considered with the absence of histone modifications at *Tex19* (*Section 4.9 & 4.10*), the data supports a model whereby promoter DNA methylation is the primary and exclusive mechanism of directing *Tex19* expression. This model fits the *in vivo* developmental expression pattern of *Tex19,* whereby expression is observed throughout the embryo until ~E7.5, when *Tex19* is *de novo* methylated and silenced, probably by Dnmt3b. In PGCs, *Tex19* either avoids this *de novo* methylation event or more likely, is specifically demethylated during an epigenetic reprogramming phase at ~E8.5 and is therefore expressed specifically in the germline (*Section 4.11*).

The data presented here strongly supports the conclusion that *Tex19* is regulated primarily by promoter DNA methylation. However when making such a defined conclusion, it is important to also consider alternative possibilities in the context of previously published data. For example, it has been reported that the polycomb (PcG) component EZH2 can serve as a recruitment platform for DNA methyltransferases (Vire et al., 2006). Thus, one possibility is that DNA methylation

at *Tex19* is a consequence of PcG activity. However in this case, my data suggests that the polycomb mark H3K27me3 is only weakly deposited at *Tex19* in silenced cell-types (*Section 4.10*). This suggests that H3K27me3 has, at best, a secondary epigenetic role at *Tex19,* a premise supported by the lack of *Tex19* activation in T-cells deficient for EZH2 and also Eed-null embryoid bodies (Su et al., 2005 & *Section 4.10*). Alternatively, an intriguing recent study noted *Tex19* exhibited a weak but significant enrichment of H3K9me3, an epigenetic mark not examined here (Yuan et al., 2009). It is therefore important to consider the possibility that this epigenetic modification might have a role in regulating *Tex19.* However, examination of the data suggests that H3K9me3 was localised significantly downstream of the *Tex19* promoter and was only weakly enriched. Moreover, despite being typically associated with a repressed chromatin structure, the H3K9me3 enrichment was present in ES cells, where *Tex19* is strongly expressed. These observations indicate that like H3K27 methylation, H3K9me3 is not a significant part of the *Tex19* regulatory mechanism. In addition to epigenetic factors, it is important to consider that transcription factors could functionally operate upstream of CpG methylation at *Tex19.* For example, it has been demonstrated that E2F6-deficient mice exhibit dysregulated expression of *Tex12* coupled with demethylation, indicating that promoter methylation is a consequence of E2F6 activity at this locus (Pohlers et al., 2005). However, Pohlers and colleagues noted that two genes identified as methylation-dependent candidates here (*Tex13* & *Mov10l1*), were not de-repressed by E2F6-deficiency. This is consistent with the premise that CpG methylation, and not E2F6, is the primary upstream mechanism regulating the candidate genes identified here, including *Tex19.*

However, to consider the broadest possible scope of alternative epigenetic or trans-acting factors that could contribute to *Tex19* regulation, I performed a comprehensive analysis of *Tex19* expression across the entire body of published microarray data deposited in the GeoProfiles database. I found *Tex19* was not dysregulated in any loss or gain of function studies of epigenetic mediators, including HDAC1, HDAC2, Sirt1 (HDAC), EZH2 and linker Histone H1 (Zupkovitz et al., 2006; Alcendor et al., 2007, Su et al., 2005, Fan et al., 2005b) or transcription factors, including; Rb, Myc,

c-Jun, Oct3/4, the MAPK pathway, Wilms-Tumour protein, Lim1 and Nanog (McCabe et al., 2005; Lawlor et al., 2006; Drosatos et al., 2007; Maekawa et al., 2005; Klattig et al., 2007; Potter et al., 2007; Loh et al., 2006). In contrast, the top hits for *Tex19* de-repression were from forced DNA demethylation studies (Vallender & Lahn, 2006, Perez-Iratxeta et al., 2005 (GDS2905), Lande-Diner et al., 2007). This analysis, and that presented in this thesis, has thus considered the widest possible range of alternative interpretations for the mechanism of *Tex19* regulation. Taken together they provide compelling evidence that *Tex19* is causally regulated exclusively by promoter DNA methylation upstream of any epigenetic or trans-acting factors. Moreover, this is consistent with a recent study which reported that promoter CpG methylation may be the exclusive epigenetic mark at ~30% of genes in ES cells (Fouse et al., 2008).

The identification of *Tex19* as a locus that is *causally* regulated by DNA methylation represents a novel example of a single-copy gene unambiguously controlled exclusively by the presence or absence of promoter CpG methylation. However, my data and others indicate that *Tex19* is unlikely to be the only target critically reliant on CpG methylation. Instead it is possible that *Tex19* belongs to a small but significant group of germline restricted methylation-dependent genes. For example, the stringent candidate screen performed here (*Chapter 3*) identified eight other germline-specific genes that could be regulated by DNA methylation. Indeed, analysis of three of these candidates demonstrated they all exhibited highly methylated promoters in somatic cells despite being CGIs (*Section 4.2*). It will be important to determine whether these genes are causally regulated exclusively by promoter CpG methylation analogously to *Tex19* or whether promoter methylation only contributes to their regulation similarly to *PiwiL2*. Here, cause and effect analysis suggested that DNA methylation was an important component of *PiwiL2* silencing but not the exclusive mechanism (*Section 4.8*). Thus, while each of the remaining candidates identified here are potentially part of a methylation-dependent subset of germline-specific genes, further causal studies are needed to determine this.

Additional putative methylation-dependent germline genes have been reported elsewhere. The foremost drawback to these studies was the lack of cause and effect evidence. This left open the possibility that promoter methylation at these loci was occurring as a consequence of transcriptional activity (Weber et al., 2007; Maatouk et al., 2006; Shen et al., 2007) or that de-repression following global demethylation was a consequence of indirect effects such as changes in transcription factor availability and/or epigenetic modifications (Tachibana et al., 2009; Oda et al., 2006; McCarrey et al., 2005; Fouse et al., 2008). However, the comprehensive cause and effect analysis of *Tex19* presented here, provides proof of principle that some or many of the reported targets could be genuine methylation-dependent genes. One strong example is *Dazl,* which was one of three genes reported to rely on promoter methylation for temporal activation in post-migratory PGCs (Maatouk et al., 2006). *Dazl* only failed to be considered as a candidate here because it was de-repressed only 4.3-fold (6-fold threshold) after 14 days recovery from 5-aza dC treatment (10.2-fold immediately after). Thus, considering *Dazl* is de-repressed in both DP and P-aza cells, effectively loses epigenetic memory after demethylation and that its methylation status *in vivo* correlates with expression (Maatouk et al., 2006), *Dazl* represents an excellent future candidate. However, that the other two genes reported by Maatouk and colleagues (2006) (*Mvh* and *Scp3)* were not de-repressed in DP or P-aza cells highlights the requirement for downstream cause and effect analyses, as performed here (*Chapter 4*), to confirm causal regulation by promoter methylation. Indeed, the example of *PiwiL2* supports the necessity for comprehensive analysis to determine the precise regulatory contribution of DNA methylation at each locus

Until comprehensive cause and effect studies are carried out on germline genes putatively regulated by DNA methylation, the number of genes that primarily utilise promoter methylation to regulate developmental and tissue specific expression will remain unknown. However, the data presented here suggest that it is possible that there *are* a significant number of germline-specific genes that rely exclusively on this mechanism. Indeed, *Tex19* may only represent the first *bona fide* example of multiple methylation-dependent germline genes.

6.2 Developmental targeting of *de novo* methylation to *Tex19* and germline-specific genes

The identification of *Tex19* and several other putative germline genes regulated by promoter CpG methylation raises several salient questions. For example, which enzyme targets methylation to CGIs during development? How are germline-specific CGI promoters, such as *Tex19,* specifically targeted for *de novo* methylation in somatic cells while other CGI promoters remain protected? And how do PGCs uniquely protect these promoters from methylation and/or direct their demethylation? To address the former question, it is notable that my data indicates that Dnmt3b is responsible for *de novo* methylation of *Tex19* in ES cells during embryoid body (EB) formation. Because embryoid body development closely parallels *in vivo* development, it is likely, though not certain, that Dnmt3b plays the prominent role in methylating *Tex19* during embryogenesis. Indeed, Dnmt3b is expressed at its peak in E6.5 mouse embryos whereas Dnmt3a is absent or at best, present at very low levels at this stage (Watanabe et al., 2002 & UniGene EST database). Analysis of developmental *Tex19* methylation suggests the *de novo* process starts at ~E6.5 and is complete by E7.5 at the latest – timings that fit well with the developmental expression of *Tex19* and importantly, with the expression of Dnmt3b, but not Dnmt3a. This observation strongly supports the conclusion that Dnmt3b is responsible for *de novo* methylation of *Tex19 in vivo.*

However whether Dnmt3b is responsible for general germline-specific CGI methylation is an intriguing question. Dnmt3a and Dnmt3b have been shown to have both distinct and overlapping functions. For example, both Dnmt3a and Dnmt3b target *Xist*, the noncoding RNA essential for female X-inactivation (Sado et al., 2004) whereas Dnmt3a is solely responsible for the establishment of the majority of mono-allelic imprinted marks in germ cells (Kaneda et al., 2004). This suggests that *de novo* methylation of germline specific CGI promoters could conceptually be a distinct biological role of a single *de novo* methyltransferase, potentially Dnmt3b. Moreover the strong expression of Dnmt3b compared to Dnmt3a during early development is consistent with Dnmt3b being the primary methyltransferase

responsible for general germline CGI *de novo* methylation (e.g. at *Tex13, PiwiL2, Dazl* etc). Dnmt3b inactivation also leads to a significantly earlier developmental phenotype than Dnmt3a inactivation (Okano et al., 1999). Arguing against Dnmt3b being responsible for developmental methylation of CGIs, it has been reported that the *de novo* methylation of *Rhox6* and *Rhox9* is dependent on both Dnmt3a and Dnmt3b (Oda et al., 2006). However, examination of this data can reconcile this apparently contradictory conclusion to the Dnmt3b-dependent model proposed here. The authors found that methylation of *Rhox6* and *Rhox9* in *Dnmt [3a 3b]*-double null ES cells could be rescued by overexpressing Dnmt3a. This suggests that Dnmt3a expressed at nonphysiological levels has the capacity to target *de novo* methylation to these loci. However, analysis of embryonic tissue indicated that *Dnmt3a*-null cells retained full methylation at *Rhox6* and *Rhox9* but crucially, that *Dnmt3b*-null embryonic cells were demethylated. This suggests that *in vivo* Dnmt3b alone methylates these loci but that when present, Dnmt3a has the capacity to target these genes. Because Dnmt3a is expressed very weakly during early development it is likely that there is a crucial reliance on Dnmt3b to methylate *Rhox6* and *Rhox9 in vivo*. This data is thus ultimately consistent with Dnmt3b being the primary enzyme that targets CGI promoters in somatic cells during development. It would be interesting to confirm germline-specific CGI methylation as a specific role for Dnmt3b analogous to the specific role of Dnmt3a in methylating imprinted loci during PGC maturation.

A curious aspect of the developmental timing of *Tex19* methylation is the relatively late onset of *de novo* methylation at the locus. During development *de novo* re-methylation of the bulk genome commences from the early blastocyst stage (~E3.5) and is largely completed by the onset of gastrulation (~E6.5) (*Fig 1.7*) (Santos et al., 2002). However, my data suggests that *Tex19* is only partially methylated by E6.5 and not fully methylated until E7.5, albeit at the latest (*Section 4.11)*. While this methylation profile is consistent with the developmental expression of *Tex19,* it is intriguing to consider why *de novo* methylation at this locus is relatively delayed compared with the bulk genome (*Fig 6.1 upper panel*). One possibility is that late-onset timing of *Tex19* methylation and silencing was evolutionarily selected for to

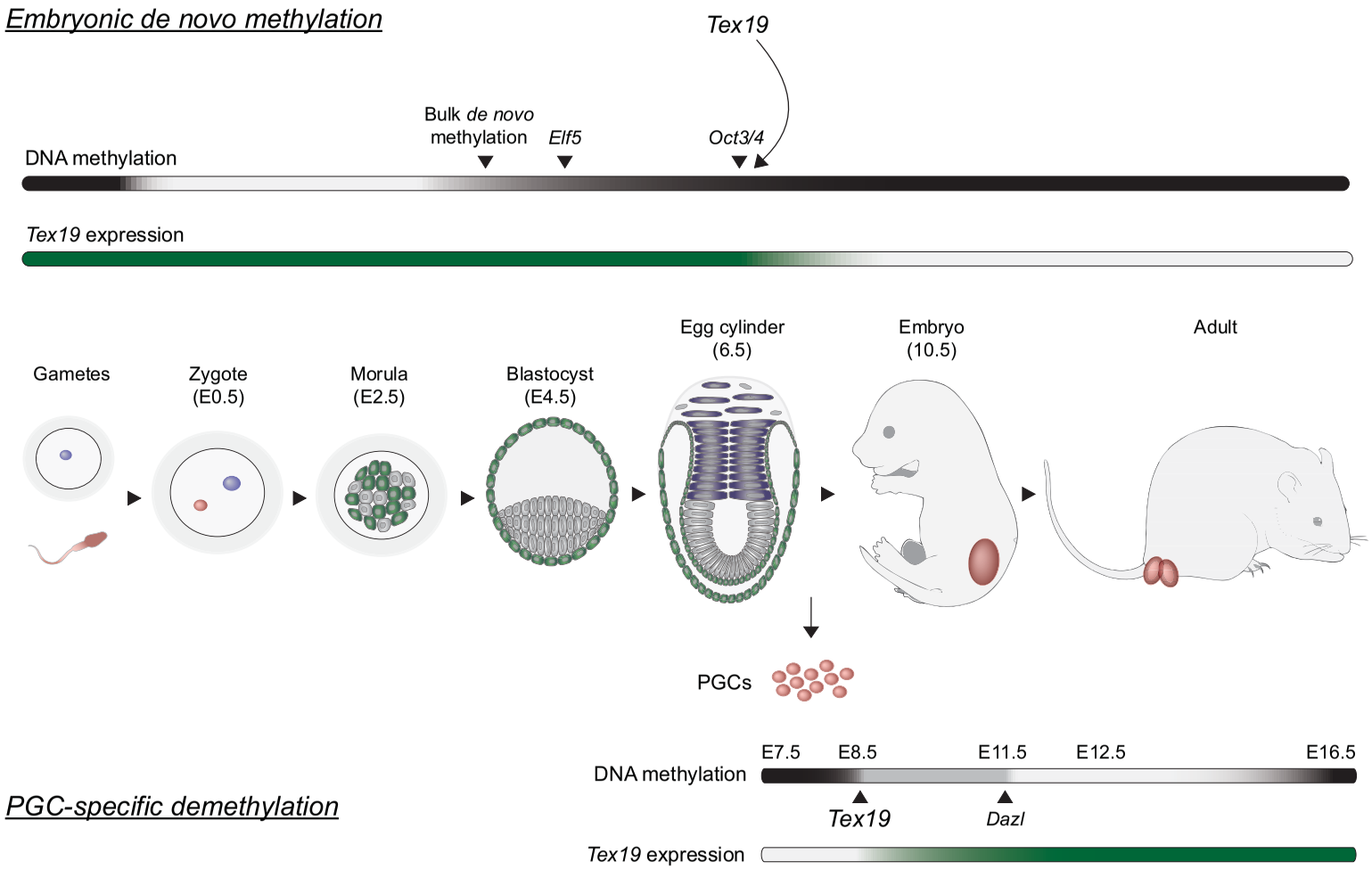


Fig 6.1. **Developmental dynamics of *Tex19* methylation and expression.** *See overleaf for description.*

***Fig 6.1*. Developmental dynamics of *Tex19* methylation and expression.** Shown is genome-wide methylation levels (black) and *Tex19* expression (green). Upper panel: During early embryonic development (zygote to E7.5) *Tex19* remains hypomethylated and expressed throughout all embryonic and extra-embryonic tissues. A delayed wave of *de novo* methylation is targeted to *Tex19* between E6.5 and E7.5 leading to stable silencing of *Tex19* expression in all somatic tissues by late E7.5. Lower panel: Germ cells are specified at ~E7.5 and undergo an initial wave of epigenetic reprogramming at ~E8.5 leading to partial DNA demethylation (*Fig 1.7)* and activation of *Tex19* and other methylation-dependent germline genes specifically in PGCs (*Fig 4.16 & Appendix 5*). A second wave of reprogramming at ~ E11.5 leads to complete erasure of DNA methylation in PGCs and activation of a second set of methylation-dependent germline-specific genes including *Dazl.*

promote the potential biological function of Tex19 (*Section 6.4*). Mechanistically, *Tex19* may be bound by specific *cis*-regulatory proteins which protect the promoter from *de novo* activity until ~E6.5. Alternatively, *Tex19* may be inherently resistant to *de novo* methylation (it is a CGI) and therefore may require a high local concentration of Dnmt3b. Indeed Dnmt3b reaches its peak expression at ~E6.5, precisely the timepoint *de novo* methylation is predicted to occur at *Tex19* (UniGene EST database; Watanabe et al., 2002). It will be interesting to investigate the developmental timing of *de novo* methylation at other germline-specific candidate genes such as *Tex13* and *Dazl* to determine if this relative delay is a general characteristic of developmentally methylated CGIs. Indeed supporting this premise, *de novo* methylation of the *Oct3/4* CGI promoter is also delayed, occurring after E6.5, thereby hinting that there may be conserved or parallel mechanisms of delayed *de novo* activity at specific CGI promoters *in vivo* (Gidekel & Bergman, 2002).

While the evidence suggests Dnmt3b is responsible for catalysing *de novo* methylation of *Tex19* at ~E6.5 *in vivo,* the mechanism that targets this activity to *Tex19* and indeed other germline-specific CGI promoters but excludes the vast majority of genomic CGIs is unknown. One possibility is that antisense mechanisms alter chromatin structure or recruit methyltransferases to germline CGIs thereby promoting *de novo* activity at these loci (Chotalia et al., 2009; Sleutaels et al., 2002). However, my analysis suggested that the antisense RNA I identified at *Tex19,* was transcribed as a consequence of *Tex19* expression and was therefore downstream of promoter methylation status at this locus (*Section 5.7*). This supports the notion that antisense mechanisms may potentially have a locus-specific role in targeting methylation to non-imprinted germline CGI genes but this is not the mechanism at *Tex19*. I therefore considered the possibility that *cis*-regulatory elements had a role in directing *de novo* methylation to *Tex19* in somatic cells. Retroelement derived sequences and tandem repeat arrays are rapidly targeted for CpG methylation in mammals (Walsh et al., 1999; Reinhart et al., 2006; Yoder et al., 1997) and also partially avoid demethylation during epigenetic reprogramming (Hajkova et al., 2002). My analysis demonstrated that the proximal *Tex19* promoter is composed of

~49% retroelement derived sequences (*Section 5.8*). Therefore it is possible that these insertions contribute to or direct *de novo* methylation to this locus. However, further analysis of both this locus and other methylated CGIs is necessary to validate this proposal.

An alternative possibility is that the inherent CpG density at *Tex19* and methylated germline CGI promoters promotes *de novo* methylation. Indeed, it is striking that all 9 of the germline-specific candidate genes identified here have promoters that are of very similar CpG density (obs/exp 0.53-0.6) (*Section 3.8*). This corresponds to promoter classifications that are on the border between ICP and HCPs. Thus, while considered as CGIs, my candidate genes and other reported genes such as *Dazl* are all at the lower limit of CGI classification (weak-CpG islands). As several genome-wide studies have reported that precisely this class of CGI promoter is preferentially *de novo* methylated during development (Weber et al., 2007; Meissner et al., 2008), it is possible that germline candidate genes including *Tex19* are targeted due to their inherent promoter characteristics or 'CpG-ness'. However, an argument against this proposal is that many genes that have equivalent CpG densities to *Tex19* and candidate loci do not become methylated during development. Therefore in this model, at least one other system in addition to CpG density must co-ordinately direct CGI methylation. Structural studies have revealed that the Dnmt3a-Dnmt3L complex methylates DNA templates with an optimal CpG spacing of 8–10bp, hinting that Dnmt3b could have a similar preference (Jia et al. 2007). Moreover, it has been reported that Dnmt3L is structurally inhibited by methylated Histone 3 lysine 4 (H3K4) (Ooi et al., 2007). Thus, the presence of unmethylated H3K4 at weak CpG islands and/or with a specific CpG periodicity could potentially account for targeting methylation to germline promoters. Interestingly, my data indicated that the core promoter of *Tex19* was depleted of histone modifications and could potentially remain nucleosome-free in some contexts. This may reflect an important aspect of targeting *de novo* methylation to this locus.

However, a recent report argues against this proposal and suggests that CGIs generally remain methylation-free due to sequence specific motifs irrespective of

CpG density or histone modifications (Straussman et al., 2009). In this model CGIs avoid *de novo* methylation during early development because they contain specific sequence motifs that confer protection. The authors argue that conceptually, methylation of CGIs is the default state unless they are protected by specific DNA sequences. Another study has additionally reported complementary sequence motifs associated with methylated CGIs (Shen et al., 2007). In spite of this, cross-referencing these motifs with *Tex19* demonstrated that sequences associated with methylated CGIs were not present at *Tex19* and conversely that the motif most highly enriched at unmethylated CGIs (CGCGC) was present adjacent to the *Tex19* TSS. Thus, this argues against sequence specificity targeting *Tex19* for developmental methylation.

The body of evidence presented here strongly supports the notion that Dnmt3b is responsible for *de novo* methylation of *Tex19* at ~E6.5 and indirectly suggests this is a general role for Dnmt3b. However, the precise mechanism that targets *Tex19* and indeed germline-restricted CGI promoters in general is not determined in this thesis. Thus, it remains speculation as to how CGI promoters are located and selected by the *de novo* methylation machinery *in vivo.*

6.3 Evolution of DNA methylation as a regulatory mechanism

While it is beyond the scope of this thesis to determine how germline-specific CGI promoters are targeted for *de novo* methylation, a key observation is that several *are* targeted for methylation (*Section 4.1*) (Meissner et al., 2009; Lister et al., 2009) . The question therefore follows, why are these germline-specific CGI promoters afforded this privileged state? In other words, why does DNA methylation preferentially regulate germline-specific genes. One answer to this question lies in the inherent mutability of methylated cytosines to thymine by deamination. In this model germ cell expressed genes would be hypomethylated in the germline and therefore avoid losing CpGs from their promoters over evolutionary time, as only germ cell DNA is transmitted to progeny (Siegfred & Cedar, 1997). In contrast, a somatic cell-specific gene regulated by CpG methylation would, by definition, be hypermethylated in the

germline. This would lead to the unstable situation whereby the promoter would have a tendency to progressively lose CpGs by deamination over successive generations (MacLean II & Wilkinson, 2005). Thus, through being impervious to evolutionary loss of CpGs, germline-specific genes have the unique advantage of being capable of preserving their ability to be regulated by promoter DNA methylation. Therefore conceptually, not only are germline-specific genes preferential targets for regulation by CpG methylation, over long evolutionary stretches they are the only class of genes able to maintain methylation-dependent regulation (with the exception of imprinted loci).

Even though germ cell-specific expression is effectively a requirement for stable methylation-dependent regulation, only a limited subset of germline-restricted genes are regulated by this mechanism for several putative reasons. Firstly, many germ cell-specific genes are associated with low CpG density promoters. As suggested by Weber et al (2007), below a certain threshold of CpG density, DNA methylation has little or no effects on transcription. Thus, a prerequisite for being regulated by DNA methylation is a promoter containing a sufficient density of CpGs to interfere with transcription. Secondly and perhaps more importantly, many genes are regulated by the co-ordinated effects of epigenetic and trans-acting factors. Therefore, if a gene required a germ cell-specific transcription factor for activation (e.g. FIGalpha or Sohlh2), promoter demethylation may be necessary, but would not be sufficient for gene expression in somatic cells (Soyal et al., 2000; Ballow et al., 2006). At such loci, DNA methylation may contribute to gene regulation but would not be the exclusive mechanism. *PiwiL2* and previously reported genes such as *Tex12, Rhox6* and *Pgk2* may fall into this class of genes whereby DNA methylation is a crucial contributor to regulation but transcription factor availability and/or additional epigenetic modifications also modulate the precise cell-type expression pattern (Pohlers et al., 2005; Oda et al., 2006; McCarrey et al., 2006). In contrast, germline-specific genes that are activated by ubiquitous transcription factors would critically rely on DNA methylation (or an alternative epigenetic mechanism) to maintain silencing in somatic cells. Only at these germline-specific loci could DNA

methylation potentially be the primary and exclusive mechanism for regulating tissue specific expression.

Thus, for a gene to be regulated by DNA methylation, it must satisfy several prerequisites. It should be (i) germline specific, (ii) activated by ubiquitous transcription factors, (iii) have a CpG-dense promoter and, (iv) not have acquired additional upstream epigenetic regulation. Additionally it must acquire tissue-specific *de novo* methylation during development. I propose the necessity for these several requirements imposes a severe limit on the number of genes that can theoretically be regulated primarily by promoter CpG methylation. This accounts for the lack of *bona fide* methylation-dependent targets currently characterised and is consistent with *Tex19* relying exclusively on DNA methylation, as it satisfies these requisites. It is noteworthy that the requirements outlined here are not *strictly* necessary (but are likely) for methylation-dependent regulation. For example a somatic cell-specific gene could conceptually be regulated by DNA methylation if methylation mediated control was established so quickly over evolutionary time that few CpGs were lost by deamination and were then maintained by strong selection. Additionally, genes targeted specifically by germline transcription factors can be temporally regulated by DNA methylation *per se* in germ cells although crucially, loss of methylation in somatic cells would not induce activation. Thus, while the requisites proposed here generally apply, there may be exceptions.

The discussion thus far has considered the technical reasoning for why germline-specific genes are regulated by DNA methylation i.e. CpG density, retroelement derived promoters, germline deamination etc. However, it is also interesting to consider why these genes are regulated by promoter methylation from an evolutionary-developmental (evo-devo) perspective. My data indicates that *Tex19* is fully methylated in embryonic cells before E7.5, which is about the same time as germ cells are specified in the primitive hindgut (Saitou et al., 2002; Ohinata et al., 2005). This suggests that nascent PGCs are methylated at *Tex19*. However by E9.5 *Tex19* is demethylated and expressed in PGCs but remains fully methylated and silenced in somatic cells at the same stage (*Section 4.11*). Therefore it is likely *Tex19*

is demethylated as a consequence of the global epigenetic reprogramming phase that occurs in PGCs at ~E8.5 (*Fig 6.1 lower panel*) (Seki et al., 2005; Seki et al., 2007; Hajkova et al., 2008). Likewise, other germline specific genes such as *Dazl* are demethylated and expressed during the second wave of PGC reprogramming at ~E11.5 (*Fig 6.1*) (*Section 4.11*) (Maatouk et al., 2006; Hajkova et al., 2008). It is interesting to consider the possibility that genes such as *Tex19* and *Dazl* are potentially expressed specifically in the germline simply as an indirect consequence of the requirement for genomic reprogramming. Here, *Tex19* and others may be expressed solely in germ cells by virtue of being (i) *de novo* methylated during early development and (ii) having a sufficient CpG density to silence transcription. That is, germline genes regulated by promoter methylation may be the consequence of their promoters exhibiting the novel combination of attracting developmental *de novo* methylation and containing a relatively high CpG density. Expression in the germline is then an inevitable outcome of passage through reprogramming germ cells (as long as appropriate transcription factors are available). In contrast the majority of remaining genes are either not *de novo* methylated (CGI genes) or do not contain a sufficient density of CpGs to interfere with transcription when they are methylated.

Once a germ cell-restricted expression pattern is established, it is likely that genes will acquire germline-specific functions. Indeed genetic deletion of *Tex19, Dazl* or *PiwiL2* each causes severe reproductive phenotypes (Ollinger et al., 2008; Ruggiu et al., 1997; Kuramochi-Miyagawa et al., 2004). It is possible these roles evolved after the acquisition germline-specific expression and thus as an indirect consequence of DNA demethylation during PGC reprogramming. To paraphrase, these genes may not be expressed in the germline because they have germ cell specific functions but rather have germ cell specific functions because they are expressed in the germline. Indeed, it is unclear whether expression of *Tex19* in somatic cells would have any negative effects, particularly given that its reported physiological role is in repressing an endogenous retrovirus, a seemingly advantageous function (Ollinger et al., 2008). Thus, the biological reasoning for silencing *Tex19* by promoter methylation in somatic cells is uncertain, supporting the concept that *Tex19* attracts methylation (or is unable to repel it) by default and is expressed in the germline as an indirect

consequence of genomic reprogramming. However, an alternative possibility is that DNA methylation evolved to regulate the timing and tissue-specific activation of key genes required for germ cell differentiation such a *Tex19* and *Dazl*. This model predicts that these genes co-opted DNA methylation as a regulatory mechanism to ensure stable and heritable expression in the germline.

It would be interesting to distinguish between these possibilities by examining the role of promoter methylation in germline-specific gene regulation in distantly related organisms. For example, if DNA methylation was found to regulate *Dazl* in salamanders (*Axolotl*), which specify their germ cells through an 'inductive' mechanism similar to mammals but not in *Xenopus*, which specify their germline through a 'preformation' germplasm mechanism, this would support the former model (Johnson et al., 2003; Lawson et al., 1999; Seydoux and Braun, 2006). That is, that DNA methylation regulates germline specific genes as an indirect consequence of the requirement for epigenetic reprogramming of germ cells. However, if both organisms were found to regulate *xDazl* and *axDazl* via DNA methylation, this would support the latter model, which predicts that DNA methylation evolved at key germline genes to direct expression in germ cells regardless of how they are specified. Interestingly, identification of multiple methylation-dependent germ cell-specific genes in either organism would suggest that the evolutionarily role of DNA methylation in regulating germline gene expression predated its role in genomic imprinting (which is mammalian specific). Thus, this would indicate DNA methylation initially arose as a gene regulatory mechanism to promote germ cell development.

6.4 Methyl-binding proteins couple DNA methylation to gene silencing

While the role of promoter CpG methylation at *Tex19* in mice seems clear, the precise mechanism through which transcriptional silencing *per se* is achieved is unclear. Indeed, several mechanisms have been proposed to underpin the relationship between DNA methylation and transcriptional silencing (Klose & Bird, 2006).

However, it is still uncertain what contribution each putative mechanism has to gene silencing at both a global and locus-specific level. In this thesis, I have demonstrated that the methyl-binding protein (MBP) Kaiso, critically couples DNA methylation to transcriptional silencing of *Tex19*. In the absence of Kaiso, *Tex19* is strongly de-repressed despite retaining promoter methylation, establishing that Kaiso interprets CpG methylation at this locus and directs downstream silencing (*Section 5.2*). Consistent with this, *Tex19* silencing can be partially re-imposed by expression of exogenous xKaiso (*Section 5.4*). The fact the *Xenopus* Kaiso can rescue repression suggest *Tex19* inherently attracts conserved Kaiso methyl-binding activity, probably through the multiple CGCG motifs present in the *Tex19* promoter. Indeed, Kaiso is recruited to *Tex19* in a methylation-dependent manner (*Section 5.5*). Moreover Kaiso is crucially required to repress *Tex19* transcription during ES cell differentiation, as *Kaiso*-null embryoid bodies fail to impose silencing of *Tex19.* Taken together, my results reveal that the methylation-dependent silencing of *Tex19* is critically mediated by recruitment of Kaiso. The observation that *Tex19* is regulated by Kaiso reinforces the conclusion that *Tex19* represents a novel and *bona fide* gene regulated exclusively by promoter CpG methylation (*Chapter 4*).

Interestingly, a key point is that because Kaiso is ubiquitously expressed throughout development and in all tissues at comparable levels (Unigene EST and BioGPS databases), silencing at *Tex19* is entirely dependent on the presence of DNA methylation. That is to say, differential promoter CpG methylation is the key variable that directs *Tex19* silencing as Kaiso is continually available. This is distinct from a previous study of MBP regulation, which reported that gene silencing in the colon is mediated by elevated MBD2 levels specifically in the colon and not changes in CpG methylation *per se* (Berger et al.,2007). While it is possible that $p120^{ctn}$-mediated depletion of the nuclear pool of Kaiso could affect silencing, this effect has not been observed in normal cells (Daniel & Reynolds, 1999; Daniel, 2006). Thus, the relatively stable expression and nuclear localisation of Kaiso allows a binary system of methylation-dependent regulation at *Tex19.* Kaiso therefore fulfils the role of a novel methylation-dependent repressor of this locus rather than effectively being a tissue-specific repressor, as for the discussed MBD2 example (Berger et al., 2007)

An intriguing question that arises from this study is precisely how does Kaiso bring about methylation-dependent repression of *Tex19*? One previous study has indicated that Kaiso is part of a multiprotein histone deacetylase complex in HeLa cells, where it interacts with N-CoR (Yoon et al., 2003). Here, it was reported that Kaiso targeted deacetylase activity to the methylated *MTA2* gene and that this was essential for gene silencing. While *MTA2* probably does not represent a physiological target of Kaiso, as it is unmethylated in normal cells (as opposed to HeLa cells), this report may hint at a potential mechanism of repression at *bona fide* Kaiso targets. Indeed, in the present study, *HDAC1*-null ES cells failed to silence *Tex19* after treatment with retinoic acid as compared to wild-type ES cells, suggesting HDAC activity is required to initiate *Tex19* silencing. Thus, one possibility is that Kaiso targets histone deacetylation to *Tex19* following *de novo* methylation. However, consistent with previous global expression analyses (Lande-Diner et al., 2006), I observed no de-repression of *Tex19* in somatic cells treated with the histone deacetylase inhibitor TSA. This therefore suggests a model whereby HDAC activity is targeted by Kaiso to initiate repression of *Tex19* but that once silenced this HDAC activity is dispensable to maintain stable silencing.

A key aspect of this model is the distinction between initiation and maintenance of *Tex19* silencing. Microinjection of unmethylated and methylated DNA in *Xenopus* oocytes and mammalian cells has demonstrated a crucial role for chromatin compaction in initiating silencing of methylated templates. Here, methylated DNA is initially highly transcribed but becomes progressively silenced coincident with assembly into condensed chromatin (Kass et al., 1997). As HDAC activity has an important role in initiating chromatin condensation, a potential mechanism for Kaiso mediated silencing of *Tex19* could be as follows: (i) *Tex19* acquires *de novo* methylation and recruits Kaiso (ii) Kaiso associated HDAC activity initiates chromatin compaction and (iii) once assembled into a condensed chromatin structure, CpG methylation and Kaiso co-operatively maintain transcriptional silencing (*Fig 6.2a & b*). Here, initiation of *Tex19* silencing would be critically dependent on DNA methylation, Kaiso and HDACs whereas maintenance would

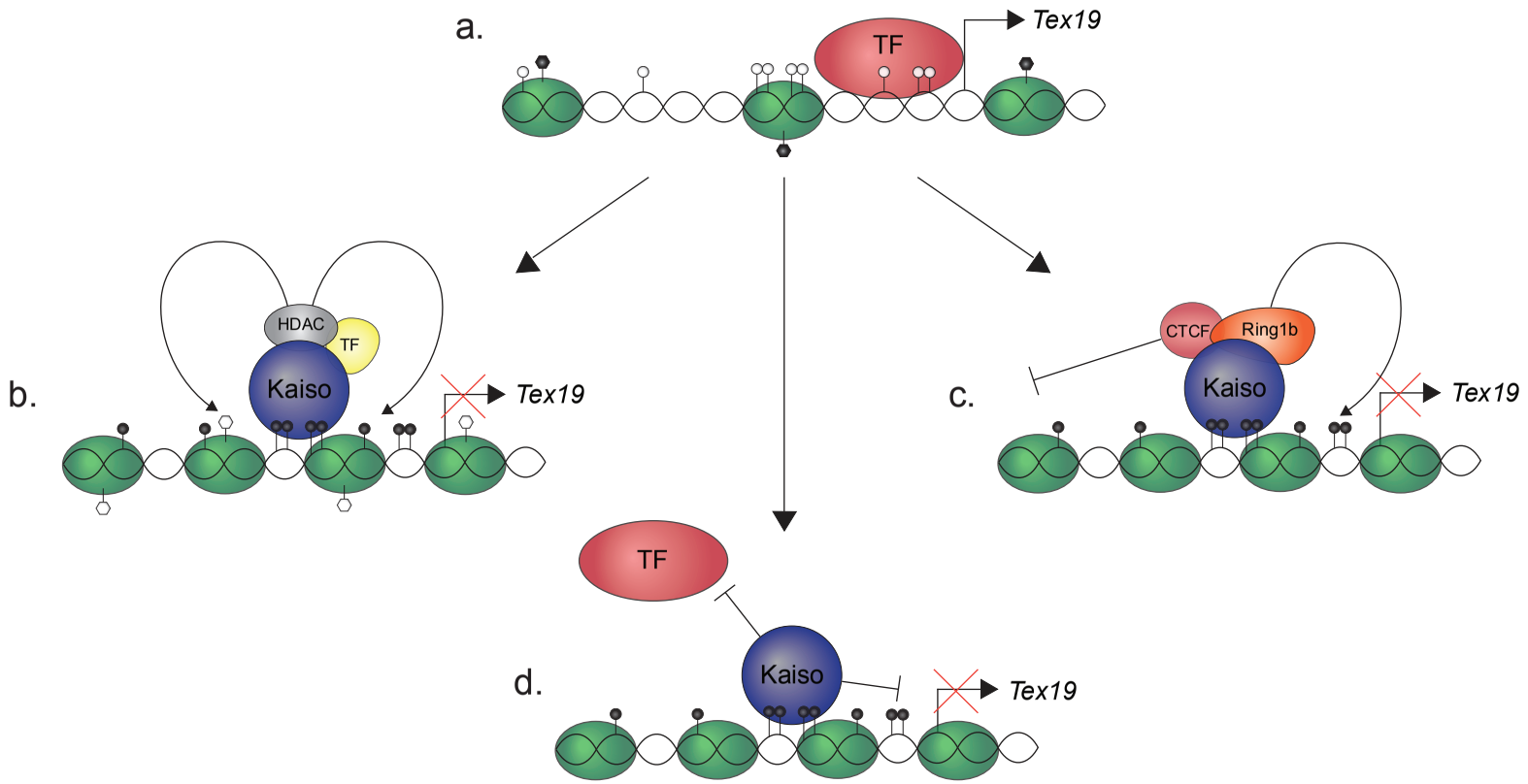


Fig 6.2. **Putative models for methylation-dependent silencing of *Tex19* by Kaiso.** a.) In the absence of promoter methylation (white circles) *Tex19* is strongly activated by ubiquitous transcription factors and is probably localised within a decondensed and acetylated (filled hexagons) genomic context. b, c, d.) Following *de novo* methylation (filled circles) Kaiso could mediate transcriptional silencing through one or several non-mutually exclusive mechanisms. b.) Kaiso could specifically recruit histone deacetylase (HDAC) activity which leads to localised deacetylation (white hexagons) and locus-compaction. c.) Kaiso could recruit alternative histone modifying proteins/complexes or chromatin insulating proteins, such as the PRC1 component and CTCF, to methylated *Tex19* to promote silencing. d.) Bound Kaiso could competitively preclude alternative transcription factors from binding the *Tex19* promoter resulting in an exclusion of activating complexes and signalling.

only rely on DNA methylation and Kaiso. Mechanistically, once *Tex19* acquires a compacted chromatin structure (methylated and deacetylated) it may directly preclude histone acetyltransferases (HATs) from gaining access, thereby negating the role of HDACs at the locus. This model provides an attractive mechanism for methylation-dependent silencing that accounts for all the experimental data generated here. However, the model predicts that *HDAC1*-null ES cells would still methylate *Tex19* during differentiation but would be unable to initiate silencing through chromatin compaction. It is therefore crucial to test this prediction to confirm the model. The observation that HDAC1 deficient ES cells fail to methylate *Tex19* after retinoic acid treatment, potentially through inhibited differentiation, would explain the failure of *Tex19* to silence in *HDAC1*-null cells (Lee et al., 2004). Moreover, when considered with the TSA data, this eventuality would suggest Kaiso directed repression at this locus was independent of HDACs. Indeed, the interaction between Kaiso and N-CoR is yet to be independently confirmed in normal mouse cells despite several attempts. It is therefore crucial for future studies to clarify the Kaiso interaction with N-CoR and the role of HDACs at *Tex19*.

There are several non-mutually exclusive alternative possibilities for the mechanism of Kaiso mediated silencing at *Tex19*. Firstly, Kaiso could act as a binding platform that may recruit multiple enzymatic activities important for maintaining chromatin structure and/or transcriptional silencing to methylated *Tex19.* In this respect, Kaiso has recently been shown to interact with the PRC1 component Ring1b and the chromatin insulator CTCF, which may contribute to co-repressor activity (Sánchez et al., 2007; Defossez et al., 2005) (*Fig 6.2c*)**.** Alternatively Kaiso may directly inhibit transcriptional initiation and/or elongation by steric hindrance. Depending on Kaiso's affinity for *Tex19* and its dynamic dissociation constant, Kaiso could dominantly prevent transcription factors gaining access to methylated but not unmethylated *Tex19* (*Fig 6.2d*). Importantly, like the Kaiso-HDAC model, these predictions are testable through yeast-two-hybrid and co-IPs or FRAP and EMSA assays, respectively. Thus, while the precise mechanism of Kaiso mediated silencing has not

been determined here, I have provided a framework for future investigations into the mechanistic basis of Kaiso activity

It is notable that, whatever the mechanism of Kaiso-mediated silencing, Kaiso is not responsible for the full programme of repression at *Tex19,* as in the absence of Kaiso *Tex19* is only partially de-repressed. Thus, although Kaiso is the critical component that ensures full silencing (i.e. turns *Tex19* completely 'off'), alternative mechanism(s) must contribute to methylation-dependent repression at this locus. Intriguingly this could be an inherent effect of the underlying DNA methylation. In this model, a deficiency of Kaiso would lead to loss of *Tex19* silencing, but maintenance of the underlying DNA methylation would sustain partial repression leading to an intermediate state of activation. DNA methylation *per se* could potentially maintain partial repression through maintaining chromatin condensation, excluding DNA binding factors, recruiting alternative MBPs or even through the direct repressive function of methyltransferases (*Fig 6.3*) (Gilbert et al., 2007; Campanero et al., 2000; Bird & Boyes, 1991; Bachman et al., 2001, Rountree et al., 2000). This model could account for the relatively weak activation of genes genome-wide in the absence of Kaiso and is consistent with *Tex19*, whereby expression in Kaiso-null cells is intermediate between hypomethylated DP cells and methylated wild-type cells (*Section 5.2 & 5.3*). It will be intriguing to determine whether *Tex19* can maintain a repressive chromatin structure in the absence of Kaiso and/or whether alternative MBPs, particularly ZBTB38 and ZTBT4, can be recruited to this locus as a redundant mechanism. In this respect, it is notable that in the present study I observed weak de-repression of *Tex19* in the absence of MBD2. Establishing the chromatin structure at *Tex19* and whether alternative MBPs are recruited could point toward the mechanistic basis of Kaiso-mediated silencing.

The observation that most genes are, at best, only partially de-repressed in the absence of Kaiso suggests a tantalising rationale for the mild phenotypes of MBP deficient mice. That is, while MBP mediated transcriptional repression may contribute to silencing at a number of loci, MBP inactivation can be compensated for by either the underlying methylation or, unlike at *Tex19,* alternative epigenetic

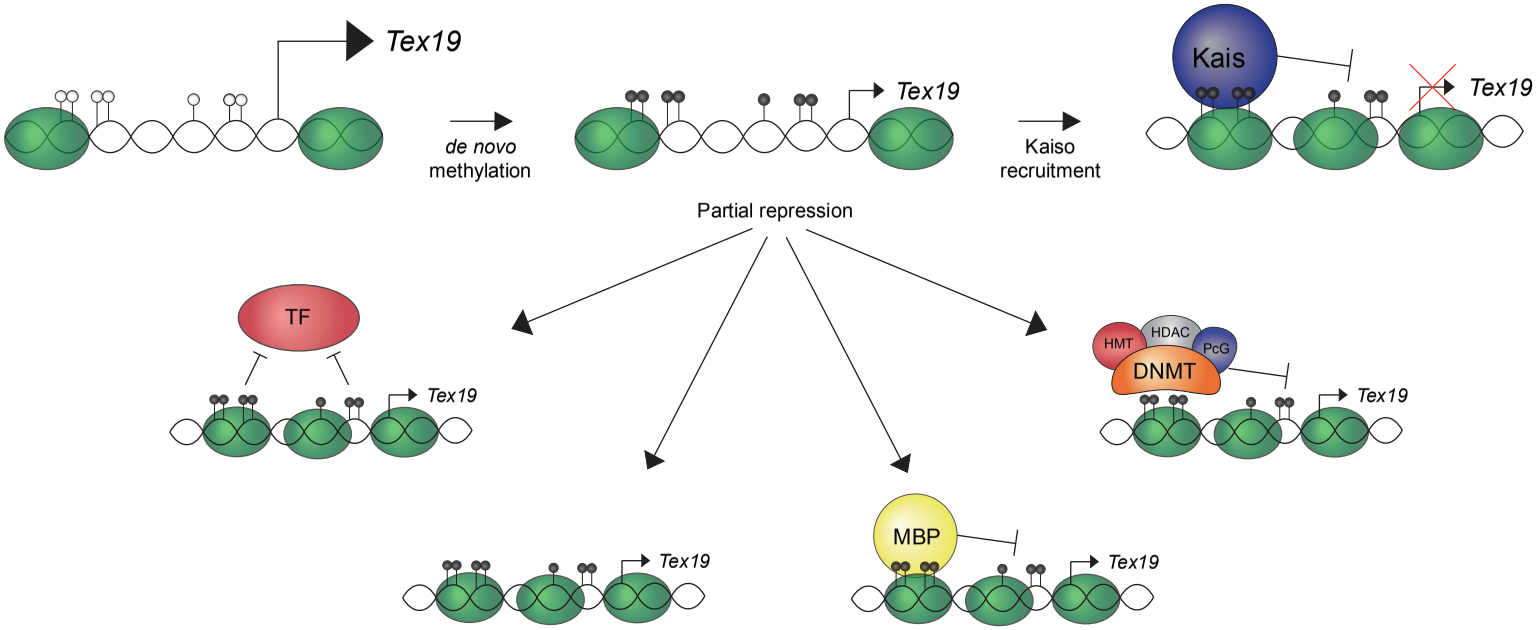


Fig 6.3. **Model of partial repression of *Tex19* by promoter methylation in the absence of Kaiso.** Unmethylated *Tex19* is strongly expressed (top left) whereas *Tex19* is silenced when Kaiso is present and its promoter is fully methylated (top right). However, an intermediate level of transcription is observed when the promoter is methylated but Kaiso is not present. This suggests that Kaiso is crucial for full silencing of *Tex19* but that DNA methylation *per se* can mediate partial repression. Shown are putative mechanisms for the maintenance of partial repression by DNA methylation when Kaiso is absent. Bottom left: Promoter methylation may directly exclude binding factors. Middle left: Methylation could directly cause locus condensation thereby precluding transcription. Middle right: Alternative methyl-CpG binding proteins such as Mbd2, or Zbtb4 could mediate partial repression. Bottom right: Alternative epigenetic modifiers may be recruited to the locus by DNA methylation to maintain partial repression.

systems (e.g. histone lysine methylation). As discussed above, DNA methylation *per se* could sustain partial silencing through maintaining repressive chromatin structure, direct exclusion or redundant MBP binding (*Fig 6.3*). Indeed this could be in addition to alterative epigenetic mechanisms at any particular locus. Thus, a consequence of self-reinforcing epigenetic regulation is that repressive networks are able to compensate for the loss of one, or even several, methyl-binding proteins. This model predicts that very few genes would be critically dependent on MBPs to maintain silencing and is thus consistent with the molecular and phenotypic data. To this extent *Tex19* may represent a rare example of a gene that is dependent on a single epigenetic regulatory pathway, namely DNA methylation. Therefore, loss of DNA methylation or one of the downstream effectors (Kaiso) leads to significant de-repression of *Tex19* compared to most loci. Interestingly, inactivation of MBPs on an already epigenetically disrupted background, such as a PcG group component or a H3K9 methyltransferase, could identify MBP targets that, on a wild-type background, would be able to compensate for MBP loss. The concept of self-reinforcing epigenetic interactions has been experimentally demonstrated between multiple epigenetic systems and it is therefore not a giant conceptual leap to include MBPs within a robust epigenetic network. Importantly, this is distinct from the reported physical interactions between for example MeCP2 and HDACs (Nan et al., 1998, Rietveld et al., 2002). Whilst this physical interaction can co-target repressive activity, the model proposed here predicts that these activities can be targeted independently and thus compensate for each other. Thus, in the absence of MeCP2, HDAC activity could still be targeted to a locus through an alternative mechanism, thus creating a robust compensatory repression system. As it stands, *Tex19* is very much the exception to this rule and thus represents the first *bona fide* MBP target gene that relies exclusively on promoter CpG methylation.

6.5 Relationship between the role and regulation of *Tex19*

Tex19 encodes a 42kDa protein that was originally identified in a subtractive hybridisation screen to identify genes transcribed in spermatogonia but not somatic cells (Wang et al., 2001). Subsequent work demonstrated that *Tex19* is expressed in

germ cells from at least E9.5 and pluripotent mES cells (Kuntz et al., 2008). The primate genome contains a single *Tex19* gene, however, a rodent-specific duplication event has produced two related genes in mice (*Tex19* & *Tex19.2*) that are separated by ~30kb and are inversely transcribed. Tex19 seems to play an important role during male germ cell maturation as deletion of *Tex19* in mice causes disrupted chromosome synapses during meiosis, and impaired spermatogenesis (Ollinger et al., 2008). Interestingly, global analysis of gene expression in *Tex19*-mutant mice identified only one transcript, the class II LTR-retrotransposon MMERVK10C, as significantly upregulated. Further analysis indicated that the elevated expression of this transcript was the likely cause of meiotic chromosome asynapses and the spermatogenesis defect in *Tex19*-null mice. It was unclear whether this is a direct or indirect consequence. However, the resemblance to other mutant phenotypes that exhibit defects in meiotic chromosome synapses as a consequence of increased retrotransposon activity in the germline is striking (Aravin et al., 2006; Bourc'his et al., 2004). This suggests that Tex19 may be a crucial component of a repressive system that regulates a specific class of LTR retrotransposon in the germline thereby promoting germ cell development. Interestingly, Tex19 is primarily localised in the cytoplasm and therefore likely regulates MMERVK10C retrotransposon levels post-transcriptionally, possibly through reducing RNA stability (Ollinger et al., 2008).

Because increased retrotransposon expression in the germline can lead to mutational transposition events that transmit to future generations, germ cells have evolved multiple mechanisms to maintain retrotransposon suppression. One key mechanism is to target retroelements for dense CpG-methylation and hence transcriptional silencing. The crucial role of this mechanism in germ cells can be seen in mice lacking genes involved in methylation-dependent transcriptional silencing. For example, deletion of *Dnmt3L, Lsh* or *PiwiL2* causes strong de-repression of LTR and LINE retrotransposons in germ cells, leading to defects in meiotic chromosomes synapses (Bourc'his t al., 2004; Da La Fuente et al., 2006; Aravin et al., 2007). Indeed, many have argued that the primary role of DNA methylation *per se* is the maintenance of transposon inactivation in germ cells - the ‘genome defence’ hypothesis (Yoder et al., 1997; Bestor & Tykco, 1996). However, one drawback of

methylation-dependent retrotransposon silencing is that PGCs must undergo phases of epigenetic reprogramming which can result in partial demethylation of retroelements (Hajkova et al., 2002). This demethylation could conceptually lead to transient expression and mutagenic genomic insertion (Walsh et al., 1998; Malik et al., 1999). To counter this, alternative mechanisms of retrotransposon suppression must be in place in germ cells.

Thus, Tex19 could represent one alternative mechanism for suppressing certain classes of retrotransposons and thereby compensating for the transcriptional effects of DNA demethylation in the germline. Strikingly, this hypothesis could account for the distinct methylation-dependent regulation of *Tex19* expression. In this intriguing model, *Tex19* expression is inherently linked to global demethylation and therefore, activation of methylation-dependent retroelements is coupled by concomitant activation of *Tex19.* Because Tex19 functions through a putative post-transcriptional mechanism of retroelement suppression, it thus acts as a compensatory mechanism for the requirement of genomic reprogramming. In effect, whenever retroelements are demethylated and expressed so is *Tex19*, ensuring that retrotransposons are constantly suppressed throughout germ cell development (*Fig 6.1*). It would be interesting to test this by examining whether MMERVK10C elements are significantly upregulated by 5-aza dC treatment of *Tex19*-null fibroblasts. Interestingly, PiwiL2 also plays an important role in suppressing retroelements by directing CpG methylation to them via piRNAs (Aravin et al., 2006). It is striking that *PiwiL2* is also highly responsive to promoter demethylation in this study and could therefore utilise global demethylation as an expression cue to re-target *de novo* methylation to retroelements. Thus, both *Tex19* and *PiwiL2* may couple their expression pattern to that of their functional target, retroelements, by utilizing the same regulatory mechanism - promoter CpG methylation.

6.6 Epigenetic memory of germline-specific genes

As part of my initial screen to identify methylation-dependent targets, I examined the transcriptional effect of 5-aza dC treatment, followed by 14 days of recovery. It was

predicted that genes that rely exclusively on CpG methylation would be unable to recover gene silencing. This analysis found an enrichment of germline specific genes including, *Tex19* and *PiwiL2,* that were still strongly expressed 14 days after 5-aza dC withdrawal. The loss of epigenetic memory at these loci is likely an inevitable consequence of being regulated primarily by a single, or at least a primary, epigenetic mechanism. It is likely *bona fide* methylation-dependent genes are unable to re-target CpG methylation, through mechanisms such as polycomb (Vire et al 2006), or maintain silencing through an alternative mechanism to DNA methylation (Feldman et al., 2006). Thus, the loss of epigenetic memory at these loci may have wide-reaching transcriptional implications. For example, many cancers exhibit progressive global hypomethylation during tumour development (Esteller, 2007). The germline-specific genes identified here, which are predicted to lose their epigenetic memory when demethylated, would therefore be likely candidates for de-repression in hypomethylated tumours. Indeed, an association between germline-specific genes and genes ectopically activated in tumours has been noted. These are termed cancer/testis antigens (CTA) and over 100 have been identified (Caballero & Chen, 2009; Scanlan et al., 2004; De Smet et al., 1996). It is possible that CTA expression only represents a consequence of global hypomethylation in cancer cells. However as expression of many germ cell specific genes is associated with pluripotent cell-types, one possibility is that aberrant loss of epigenetic memory of CTAs contributes to tumour plasticity and progression. The identification of several germline restricted methylation-dependent genes here may lead to further identification of CTAs. Moreover, this would strengthen the proposal that CTAs are de-repressed in cancers primarily due to promoter hypomethylation (De Smet et al., 1999). Indeed, genes such as *Tex19* could potentially act as biomarkers to flag epigenetically disrupted cells for clinical diagnosis or immunotherapy.

6.7 Future directions

This thesis has attempted to answer the question of whether any physiological target genes are *causally* regulated by promoter CpG methylation, and if so, by what mechanism is transcriptional silencing achieved. Here, while I have identified a

novel *bona fide* MBP- and methylation-dependent target, *Tex19,* many questions remain. One question of particular import is whether *Tex19* is a representative paradigm of multiple genes in the context of methylation-dependent and MBP-dependent regulation, or whether it is a highly specific example of this regulatory mechanism? Thus, future studies could investigate the causal role for promoter methylation at the remaining candidate genes identified here and those genes de-repressed in MBP mutant cells. Moreover as a conclusive experiment to validate the essential and causal role of CpG methylation at these loci and *Tex19*, the generation of rescued DP cell lines expressing ectopic Dnmt1 and Dnmt3b could demonstrate *de novo* promoter re-methylation imposes gene silencing. It would also be interesting to delineate the precise factors involved in *Tex19* regulation. Here EMSA assays could identify factors that bind either methylated or non-methylated *Tex19* template. This could indicate potential mechanisms for targeting *de novo* methylation to *Tex19* and hint at the key factors which activate expression of the unmethylated promoter. Indeed, we are currently generating transgenic mice expressing GFP driven by minimal or extended *Tex19* promoters to dissect the key *cis* elements required for germline specific expression and somatic cell-specific CpG methylation of *Tex19*. Finally, it would be interesting to consider the possibility that the newly discovered DNA modification 5-hydroxymethylcytosine (hmC) could have a role at *Tex19* (Tahaliani et al., 2009; Kriaucionis & Heintz, 2009). As this modification is experimentally indistinguishable from DNA methylation by bisulphite sequencing (and restriction digestion), it is possible that the observed CpG methylation profiles at *Tex19* could be entirely or in part 5'hydroxymethylcytosine. However, given the predicted low levels of this modification in somatic cells this seems an intriguing but unlikely possibility.

Appendices

Appendix 1 – Primer Sequences

Transcript	Primer Sequence	Ta	Product Size (bp)
18S rRNA	GATCCATTGGAGGGCAAGTCT CCAAGATCCAACTACGAGCTTTTT	56ºC	103
Ant4	ATGTCGAACGAATCCTCCAAGA AGCTTCACACGCTCGATGG	56ºC	140
Apobec4	ACTGCTGCATCAGCAAAATG GGCCACCACATATTGGACTC	52ºC	183
Atp5b	GGTTCATCCTGCCAGAGACTA AATCCCTCATCGAACTGGACG	56ºC	120
Beta Actin	GAAATCGTGCGTGACATCAAAG TGTAGTTTCATGGATGCCACAG	56ºC	216
B-Globin	GCACCTGACTGATGCTGAGAA TTCATCGGCGTTCACCTTTCC	54ºC	164
Crip1	AAGTGCGACAAGGAGGTGTAT AGAGGTCAGTGTCTTTCCACATT	56ºC	105
Dazl	TCTTTGCCAGATATGGCTCAGT CTTCTGCACATCCACGTCATTA	56ºC	110
Dnmt1	GCTACGAGGAGAACCACCAG GTTCCCGCTGTTACCTCTTC	54ºC	209
Dnmt1-3'	AAGAATGGTGTTGTCTACCGAC CATCCAGGTTGCTCCCCTTG	56ºC	178
Dnmt3a	TACATCAGCAAACGGAAACG GCTGCTTTGGTAGCATTCTTG	56ºC	245
Dnmt3b	CAGCTTGGAGAGGCAAAGAG TGGCTCAAGTCAACTGATGG	56ºC	216
Dnmt3L	GCTCTAAGACCCTTGAAACCTTG GTCGGTTCACTTTGACTTCGTA	56ºC	213
Dppa2	TCAACGAGAACCAATCTGAGGA GCGTAGCGTAGTCTGTGTTTG	56ºC	102
Dppa4	AGTCAACCTAGCACGGCTC TCCTGGCGTCTCAGTGTCT	56ºC	120
Eed	TCTTGGGCGATTTGATTACAG GCCACATTTATGATGGGTCAG	58ºC	175
Epiregulin	TCCGAGGATAACTGTACCGC CTCTCATGTCCACCAGGTAGAT	55ºC	136
Fshr	GAATCCGTGGAGGTTTTCG CACCTTGCTATCTTGGCAG	56ºC	171
Fthl17	TACTTTGACCGTGATGACGTG AGTTTTGCTCCAGGAAATGGC	56ºC	292
Gata 6	CTCAGGGGTAGGGGCATCA GAGGACAGACTGACACCTATGTA	56ºC	114
GCNA	CGCAACGGTTTCTGTCAGGAT GTTCAGCTCGATCATCTGGGA	56ºC	225
Gro1	CCCGCTCGCTTCTCTGTG AAGGGAGCTTCAGGGTCAAG	56ºC	232
Gapdh	ACCCAGAAGACTGTGGATGG GGTCCTCAGTGTAGCCCAAG	56ºC	289
Hdac 1	CCATGCAAAGAAGTCTGAAGC GTCTCGCAGTGGGTAGTTCAC	56ºC	271

Iap	TTGTGGCCAGAATGACAGAG GAGCGGTTCTGAGATTGGAG	56ºC	229
Iap – LTR	TTGATAGTTGTGTTTTAAGTGGTAAATAAA AAAACACCACAAACCAAAATCTTCTAC	58ºC	-
Kaiso	TGGTCCCTCAGGCTGATAAC TGCATAAACCTTGCAACCAG	56ºC	242
Krt8	GACCAAGTGGAGCCTGTTG ACGCTGTTGGATCTCATCCTC	56ºC	199
Mael	CGAGGATTTCGATTCCATTGCC GGCTCTATCATCAGACTTGCAGT	56ºC	171
Mbd1	GACTTCTGCTGCGACAAGC GAAGCAGGCCTCTTTTGATG	56ºC	182
Mbd2	CTCAGTGCTGGCAAGATGTC TCTGATTGAGGGGGTCATTC	56ºC	186
Mbd3	GCTGGGAAAGGGAAGAAGTG AAGTCGAAGGTGCTGAGGTC	56ºC	151
Mbd4	TGTGGATGGGAAAGAGTTGTG GGATTGATGCTCCCTTTCG	56ºC	197
MeCP2	CCGGGGACCTATGTATGATG AGGGTCCAAGGAGGTGTCTC	56ºC	187
Mov10L1	TTCCCTCTATGCAGGTGACAA AAGTGCATAGTGACACCGTCT	56ºC	208
Mvh	TTGGTTGATCAGTTCTCGAG CCAAAAGTGACATATATACCC	56ºC	224
Oct3/4	CAAATCGGAGACCCTGGTG AGCCTCATACTCTTCTCGTTGG	56ºC	238
p57Kip2	CAAGAGAACTGCGCAGGAG CCCAGAGTTCTTCCATCGTC	56ºC	92
Nanog	AGGGTCTGCTACTGAGATGCTCTG CAACCACTGGTTTTTCTGCCACCG	56ºC	364
Nxf2	CTGAACTGTTGTCCTTGAACT AAGGAACTGACAAGGAGAAGC	56ºC	657
Pea3	CGGAGGATGAAAGGCGGATAC TCTTGGAAGTGACTGAGGTCC	56ºC	179
PiwiL2	CATTATGGTCAAGTATCTGTT AGAGGTTGGCGAGGAATAAGG	56ºC	241
PiwiL4	ACTCCCAAACTCCGAGTCACA GGCCCGTCCACTCATGTTC	56ºC	116
Pramel1	ACTCCCATGACTTCTGTCAAT GGGAACTATATCTCCATGCCT	56ºC	448
Rbmy	AACCGAAGTAACATATACTCA ATCTGCTTTCTCCACGACCTC	56ºC	209
Rex1	CACCGACAACATGAATGAACAAAAA CAATCTGTCTCCACCTTCAGCATTT	56ºC	893
Rnh2	CCTTAAAGCACTTGTGGTAAA CCTTGCAGTCATCTGGTGAAA	56ºC	218
Sall1	CACCATGTCACGGAGGAAGCAAGCGAAGC TTACAAGGGGTTGGCAGATGTTCGTAAA	60ºC	856
Scyup	CCAGCTCTCTCTTCCTCCAC CAGCTTCTTTGGGACACCTG	56ºC	214
Slpi	GGCCTTTTACCTTTCACGGTG TACGGCATTGTGGCTTCTCAA	56ºC	148
Sox2	TAGAGCTAGACTCCGGGCGATGA TTGCCTTAAACAAGACCACGAAA	56ºC	296

Stella	CACCATGGAGGAACCATCAGAGAAAGTC CTAATTCTTCCCGATTTTCGCATTCT	58ºC	457
Sycp1	TGAGGGGAAGCTCACGGTT CGAACAGTGTGAAGGGCTTTTG	56ºC	125
Stk31	GGAAATGACCTTTCAGATGCTATGCA CTGCAGAGATGCCTCTGTT	56ºC	408
Taf7l	TCCGCTGGGAAGTCGTTGATG GTGGTTCCACCCAGTCTTCAT	56ºC	620
Tex11	TATCAGATTCCCTGGAACTGG GCACCCTCAAAACAAGCTATG	56ºC	208
Tex12	AAGAGAATTGGAGCCTCAGGT TATAATGTGCCAAATATTTGACCCTC	56ºC	489
Tex13	TTTTGGCCCACACTAAACTCG TGTAGTCTCGCACAACTCTCA	56ºC	108
Tex14	CTCATGCTCTATACTGGAAGCTG CGGATTCTACTCCGATTCCTCTT	56ºC	185
Tex15	TTCAAACCTAACAGCAGGAAA GACTGTGGAGGTATATTCCTG	54ºC	411
Tex18	GATCATTGCTTCAGGCTACCA CTTCACTTAAAAGGAGGCAAA	56ºC	453
Tex19	AAAATGGGCCACCCACATCTC CCACTGGCCCTTGGACCAGAC	56ºC	184
Tex19 3'	CGTGTCAGTGTTCAGTGTTTG ATGACAGTAAGGTCAACTAGTGC	56ºC	140
Tex19.2	TGGCTCATCCCTCCTTTGTC CAGCATGTAGCAAATGGCGTC	56ºC	198
Tex19.2 3'	CGGAAGGTTAGACTCAGCTTC AACTCTGAATCCAGGACTCAC	55ºC	200
Trap1a	CGGTGATGGGAATAGGTGCAA TGGAGCGCCAGAAAACTTGT	56ºC	111

Table A1.1. **RT-PCR primer sequences and cycling conditions.** Shown is information for all RT-PCR and qRT-PCR primers used within this thesis. Primer sequences are shown 5' to 3' with the forward primer above the reverse primer. *Ta* refers to the annealling temperature used during thermal cycling and product size is the expected size in base pairs (bp) of the amplified PCR product.

miRNA	Primer Sequence	Ta	Product size (bp)
miR 469	CCTCTTTCATTGATCTTGG	54ºC	ND
miR 17	TGATTAGTTCAGAATCTGGT	54ºC	ND
miR 25	GAGAGAAGTAGCAGTTACT	54ºC	ND
miR T4	GTAGCTCACACCTGTAAT	54ºC	ND
miR t3	TATGATTTTGAGCTCATGTAA	54ºC	ND
snRNA U6	TTCGGCAGCACATATAC	55ºC	ND

Table A1.2. **RT-PCR primer sequences and cycling conditions for miRNA amplification.** The forward primer sequence for each miRNA is shown 5' to 3'. A universal reverse primer was used in each reaction (*Section 2.4.4*). *Ta* refers to the annealling temperature used during thermal cycling and product size is the expected size in base pairs (bp) of the amplified PCR product.

Gene Locus	Primer Sequence	CpG's	Ta	Product Size
Dazl	GAGGTAATGATTTGAATAAA ATAAAAAAAAAACCCACRACCAC	33	50ºC	405
Dazl Nested	AGGTAATGATTTGAATAAAT AAAAAAACCAAAAAACCCAC	15	50ºC	240
Dnmt3L	GGTTATGAATAGGAAGGATTATT CAAAAAAATTAAAAAAAAAATATACCC	11	54ºC	627
Dnmt3L Nested	TATTTTTTAAGAGGTTGGTAGGT AACTAATTAAAAAAAACCTT		52ºC	527
Piwil2	GTTTGAGAGTAATTTTTATATAG AAATCTAATACCACTAAACC	24	50ºC	547
Piwil2 Nested	AGGTTTATTTTAAGAGGT TCCTTTCCCTCCTATTCCAA	16	52ºC	346
Tex13	GATAAATTATTATATTTGGGGGT AACCTCACCTCTCTAAAAACTA	15	52ºC	443
Tex13 Nested	TGGGTTTTAATAAATAGTTG AAAATAAACAATCCTAAAACC	13	51ºC	293
Tex19-Prox	GGAAGTATATAGGGTATTTA AAAATATTAAAACTCTAACCCCC	26	50ºC	526
Tex19-Prox Nested	GTTTTTTGTTGTTTTGGGAT CTATTCACTCTAAACAAACT	17	52ºC	373
Tex19-Distal	GGTTTTGTTTTTTGTTGTTG CATTTACATATCTCCCATAAAATC	20	52ºC	617
Tex19-Distal Nested	TTATTAAAGAGATAGGGAAGAAG ATCCCAAAACAACAAAAAAC	15	52ºC	273
Tex19.2	AAATTTTTGTGTGGTTAAGGTTG CAACAAAACCTTATAAAAAATCAAC	14	50ºC	535
Tex19.2 Nested	TTAAAGAGTTTGAGAATAAAAAG AACCCCAAAACAACAAAAAAC	13	52ºC	356

Table A1.3. **Bisulphite primer sequences and cycling conditions.**. Primer sequences are shown 5' to 3' with the forward primer above the reverse primer. Where necessary a second round of PCR with nested primer sets was used to improve specificity and yield. *CpG's* refers to the number of CpG dinucleotides present within the PCR product of each primer pair. *Ta* refers to the annealling temperature used during thermal cycling and product size is the expected size in base pairs (bp) of the amplified PCR product.

Target	Primer Sequence	Ta	Product Size (bp)
Tex19-GSP1	TGACTCTGACAAGTATTCC	48ºC	N/A
Tex19-GSP2	CTCTTGCCAGTCTCCCATCTC	56ºC	411
Tex19-GSP3	TGGTACAGCCATGCCTCATAG	57ºC	314

Table A1.4. **RACE primer sequences and cycling conditions.** Primer sequences for the foward are shown 5' to 3'. A specific tagged reverse primer was used to pair each forward primer (*Section 2.6.7*). *Ta* refers to the annealling temperature used during thermal cycling and product size is the expected size in base pairs (bp) of the final PCR product.

Appendix 2 – Genomic Methylation levels in P & DP cells

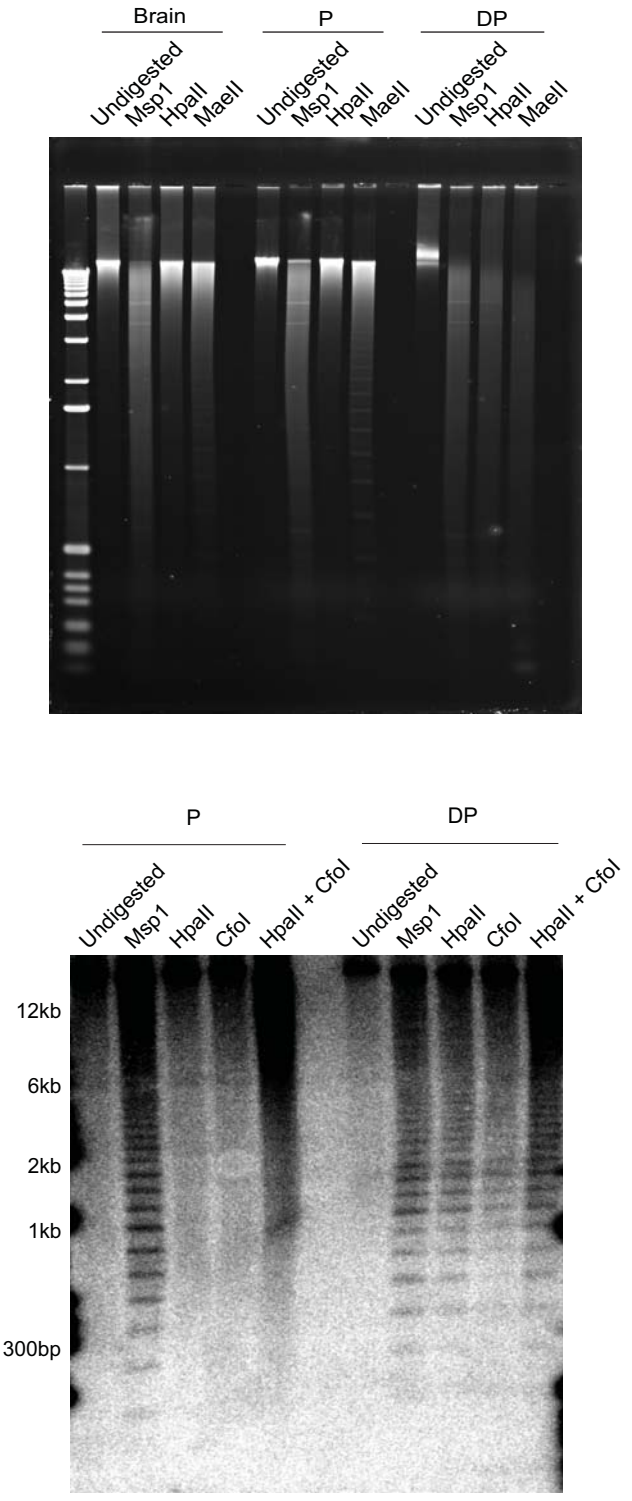
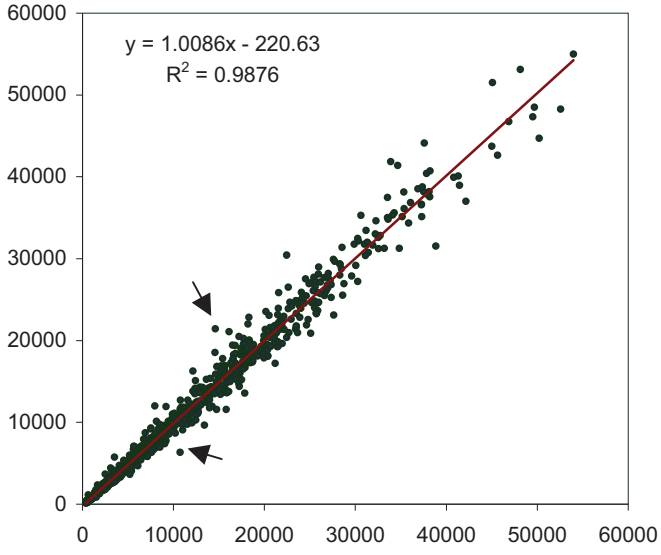


Figure A2. **Genomic methylation levels in cell lines and in vivo.** Upper panel: Genmoic DNA from mouse brain, P-MEFs or DP MEFS was digested with MspI (methyl-sensitive) or its methyl-sensitive isoschizomer HpaII and also MaeII. Digestion by HpaII or MaeII indicates hypomethylation. Lower panel: Southern blotfor mouse minor satellite following genmic DNA digestion with the indicated enzyme



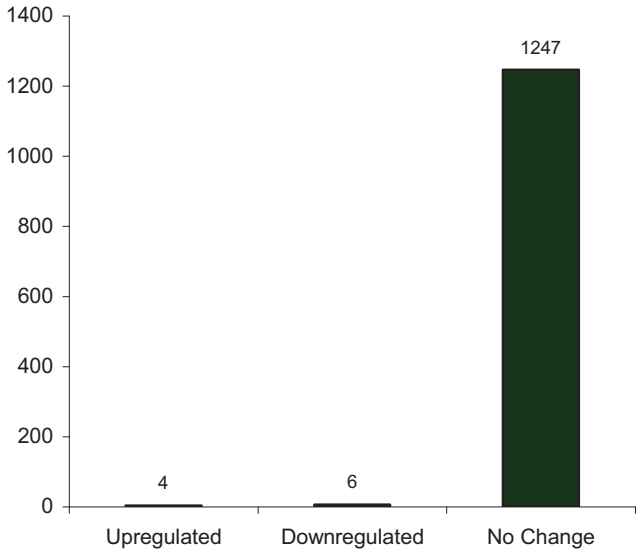


Fig A3. **miRNA expression array analysis.** The expression of diverse miRNA species in hypomethylated DP cells as compared to P cells was ascertained by through array analysis. Expression of >99% of miRNA was unchanged between the samples. Only 4 miRNAs were upregulated (>2-fold) and 4 downregulated (>2-fold). This data indicates that a globally demethylated environment is not sufficient for robust miRNA de-regulation.

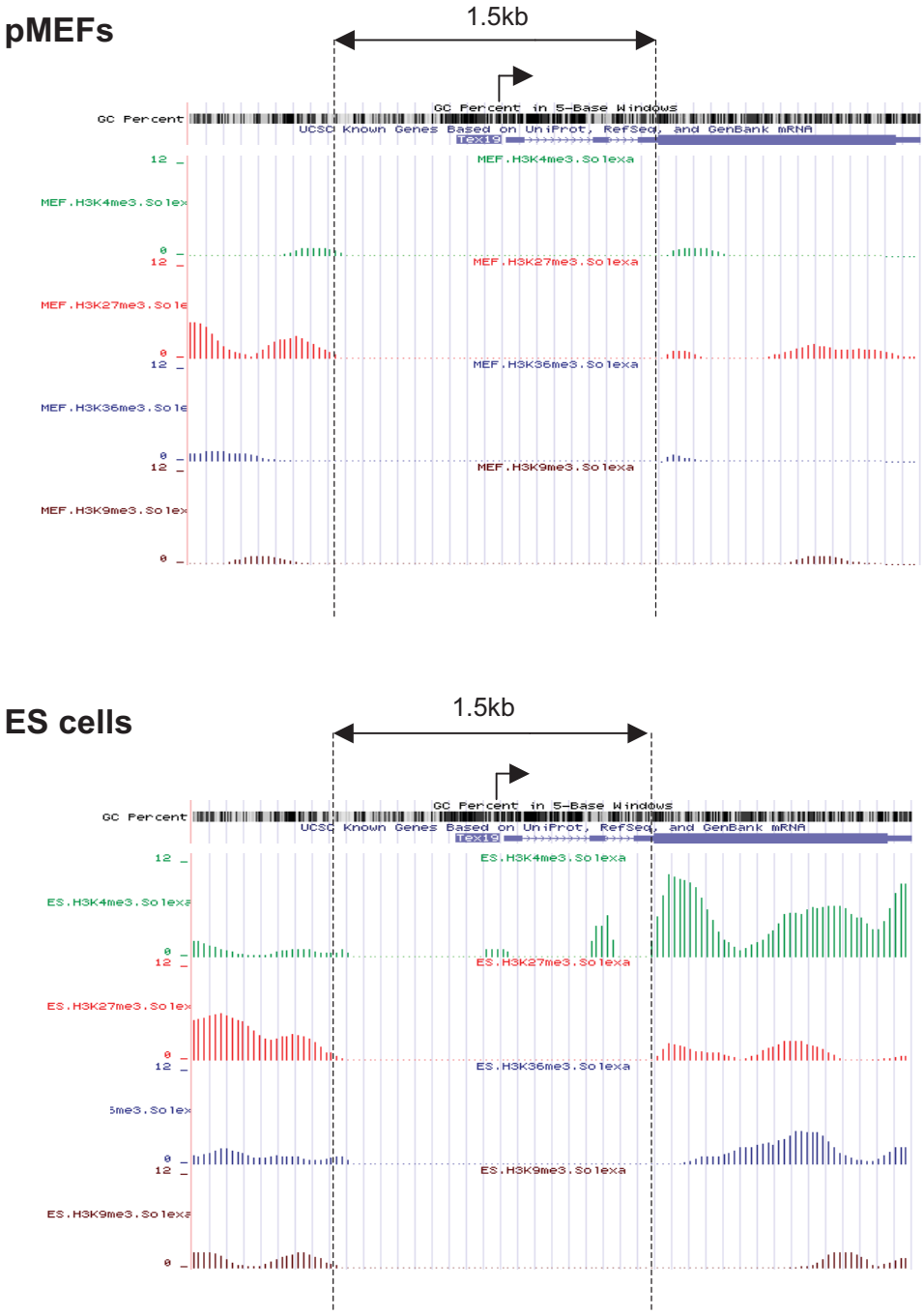


Fig A4. **Histone modification profile of Tex19 in ES cells and pMEFs derived from the Mikkelsen dataset.** Data was acquired from the Braod Chip-Seq database (http://www.broadinstitute.org/science/projects/epigenomics/chip-seq-data) using the UCSC browser. The Tex19 TSS is shown by the arrow with the modification profile of H3K4me3 (green), H3K27me3 (red), H3K36me3 (blue) and H3K9me3 (purple) shown below. No significant enrichment for any histone modification was shown 1.5kb either side of TSS in non-expressing (pMEF) or expressing (ES) cell types indicating tha histone modifications are downstream of expression changes at this locus.

Appendix 5 - PGC Methylation and Expression Analysis

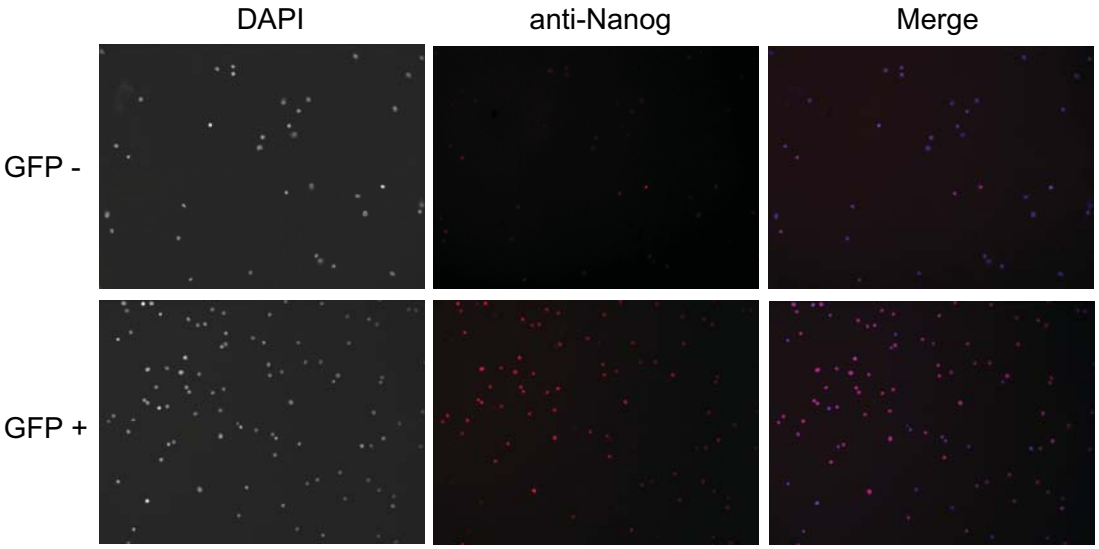


Fig A5.1. **PGC population purity.** FACS sorted PGCs and somatic flow-through were fixed and stained for endogenous Nanog expression using a monoclonal antibody (red). Nanog expression was detected in >90% of PGC and <5% somatic controls. DAPI stains nuclei.

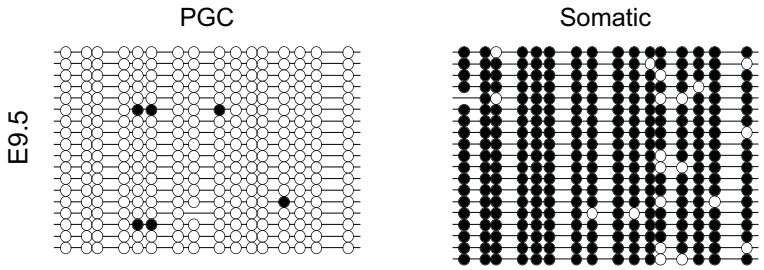


Fig A5.2. Tex19 **methylation in PGCs and somatic cells at E9.5 .** *Tex19* is hypomethylated at E9.5 in PGCs. This confirms that the hypomethylated state of *Tex19* at E10.5 (*Fig 4.2*) is not due to prmature genomic reprogramming but due to an early PGC-specific demethylation event.

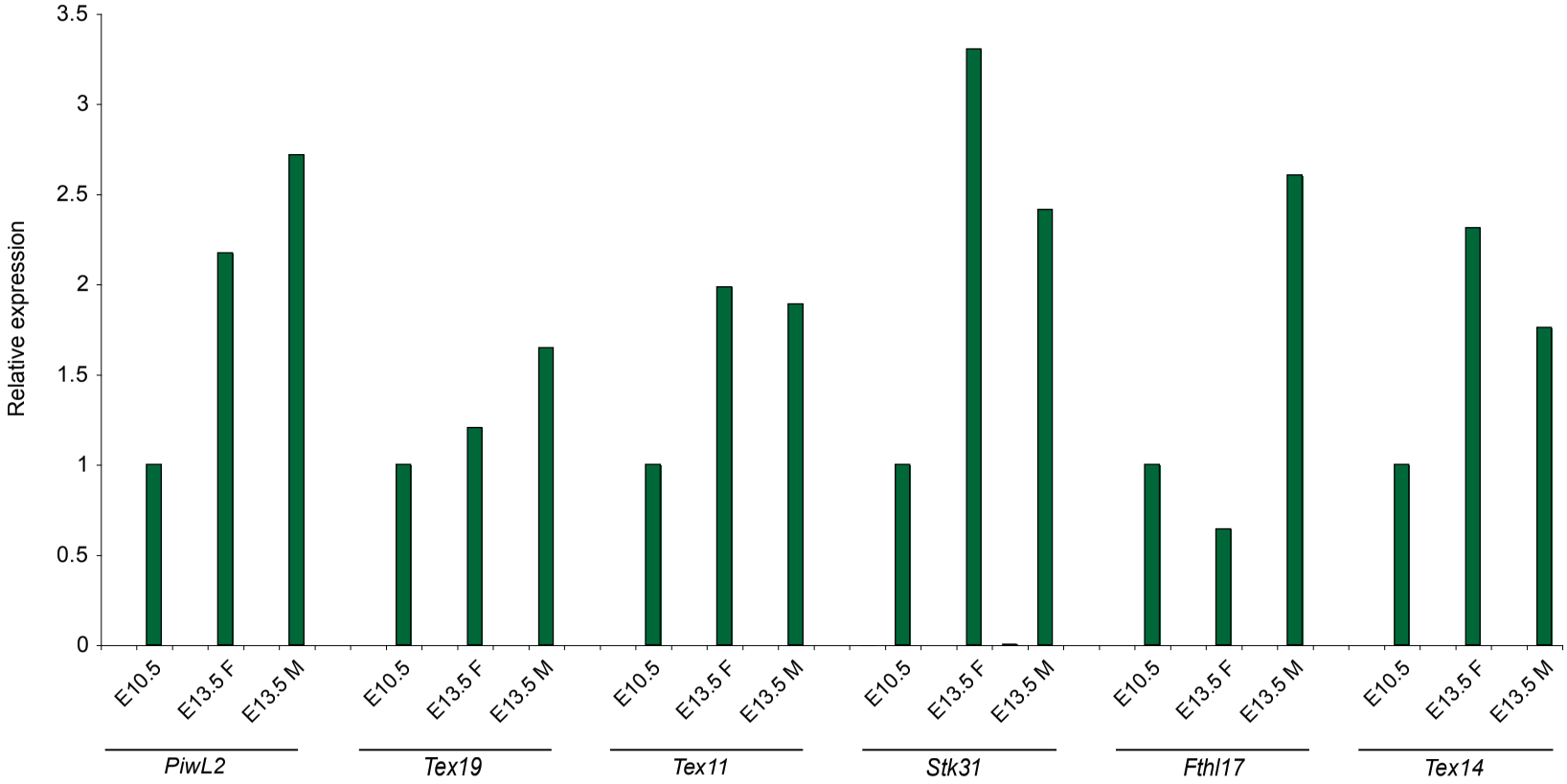


Fig A5.3. **Germline specific genes which show mininal changes in expression in PGC between E105 and E13.5.** qRT-PCR expression analysis of FACS sorted PGCs from the pre- and post- migratory stage. Shown are genes (including *Tex19* and *PiwiL2*) which did not significantly change expression during these stages. It would be predicted that these genes are already hypomethylated and expressed prior to the epigenetic re-programming event at ~E11.5. All expression is normalised to 18S rRNA and is shown relative to E10.5, which is set to 1 for each gene. Error bars represent propergated S.E.M.

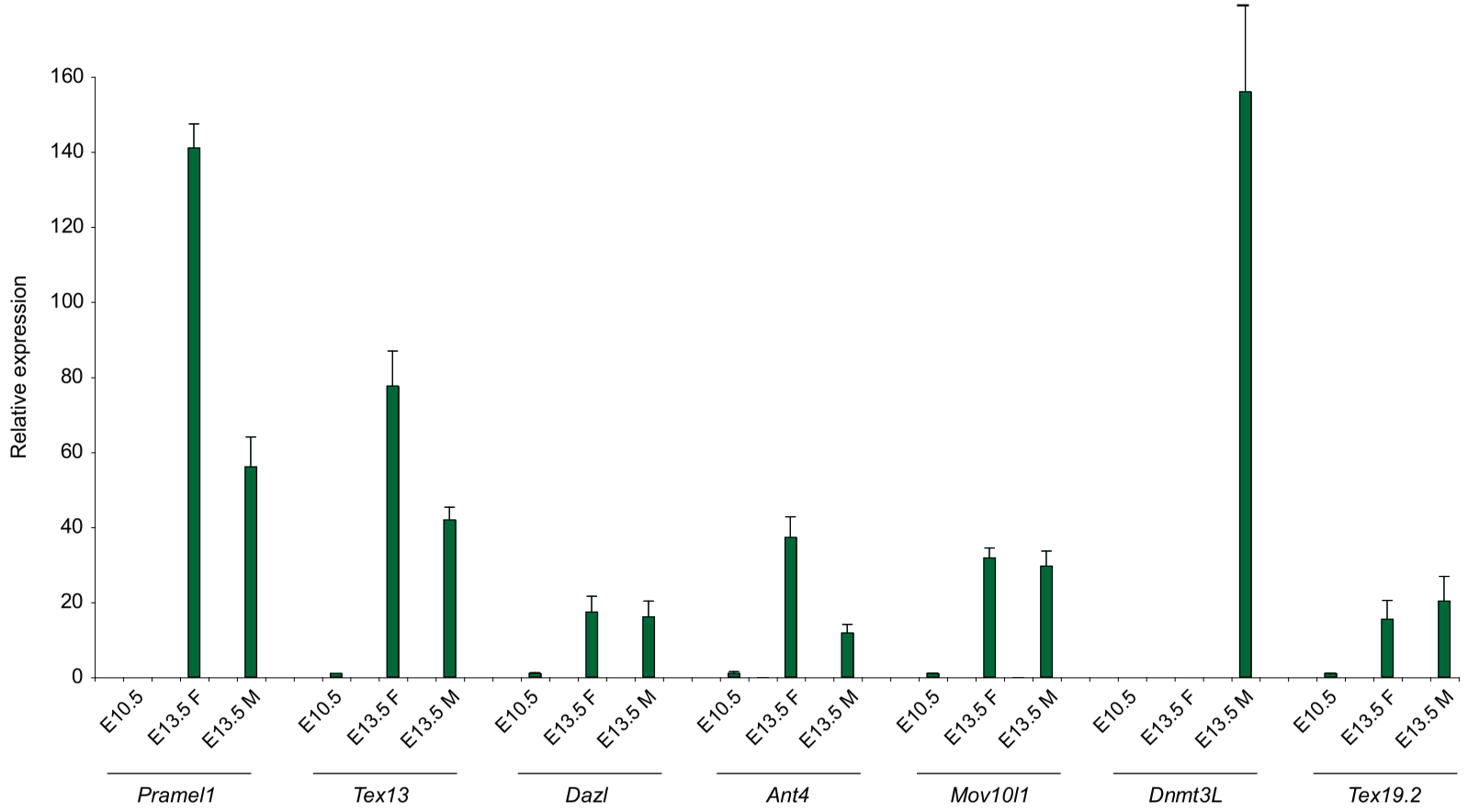


Fig A5.4. **Germline specific genes which show significant changes in expression in PGC between E105 and E13.5.** qRT-PCR expression analysis of FACS sorted PGCs from the pre- and post- migratory stage. Shown are genes which significantly changed expression during these stages. It would be predicted that these genes are hypermethylated and silenced prior to the epigenetic re-programming event at ~E11.5, at which point they are activated. All expression is normalised to 18S rRNA and is shown relative to E10.5, which is set to 1 for each gene. Error bars represent propergated S.E.M.

Bibliography

Alcendor, R. R., Gao, S. M., Zhai, P. Y., Zablocki, D., Holle, E., Yu, X. Z., Tian, B., Wagner, T., Vatner, S. F., and Sadoshima, J. (2007) Sirt1 regulates aging and resistance to oxidative stress in the heart. *Circulation Research* **100,** 1512-1521.

Allan, J., Hartman, P. G., Crane-Robinson, C., and Aviles, F. X. (1980) The structure of histone H1 and its location in chromatin. *Nature* **288,** 675-679.

Allfrey, V. G. & Mirsky, A. E. (1964) Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. *Science* **144,** 559.

Almouzni, G., Khochbin, S., Dimitrov, S., and Wolffe, A. P. (1994) Histone acetylation influences both gene expression and development of *Xenopus laevis*. *Dev. Biol.* . **165,** 654-669.

Anderson, R. M., Bosch, J. A., Goll, M. G., Hesselson, D., Dong, P. D. S., Shin, D. H., Chi, N. C., Shin, C. H., Schlegel, A., Halpern, M., and Stainier, D. Y. R. (2009) Loss of Dnmt1 catalytic activity reveals multiple roles for DNA methylation during pancreas development and regeneration. *Dev. Biol.* **334,** 213-223.

Antequera, F., Tamame, M., Villanueva, J. R., and Santos, T. (1984) DNA Methylation in the Fungi. *J Biol Chem.* **259,** 8033-8036.

Antequera, F. and Bird, A. (1999) CpG islands as genomic footprints of promoters that are associated with replication origins. *Current Biology* **9,** R661-R667.

Aoi, T., Yae, K., Nakagawa, M., Ichisaka, T., Okita, K., Takahashi, K., Chiba, T., and Yamanaka, S. (2008) Generation of pluripotent stem cells from adult mouse liver and stomach cells. *Science* **321,** 699-702.

Araujo, F. D., Croteau, S., Slack, A. D., Milutinovic, S., Bigey, P., Price, G. B., Zannis-Hajopoulos, M., and Szyf, M. (2001) The DNMT1 target recognition domain resides in the N terminus. *J. Biol. Chem.* **276,** 6930-6936.

Aravin, A., Gaidatzis, D., Pfeffer, S., Lagos-Quintana, M., Landgraf, P., Iovino, N., Morris, P., Brownstein, M. J., Kuramochi-Miyagawa, S., Nakano, T., Chien, M., Russo, J. J., Ju, J., Sheridan, R., Sander, C., Zavolan, M., and Tuschl, T. (2006) A novel class of small RNAs bind to MILI protein in mouse testes. *Nature* **442,** 203-207.

Ariel, M., Mccarrey, J., and Cedar, H. (1991) Methylation Patterns of Testis-Specific Genes. *Proc. Natl. Acad. Sci. U.S.A.* **88,** 2317-2321.

Arpanahi, A., Brinkworth, M., Iles, D., Krawetz, S. A., Paradowska, A., Platts, A. E., Saida, M., Steger, K., Tedder, P., and Miller, D. (2009) Endonuclease-sensitive regions of human spermatozoal chromatin are highly enriched in promoter and CTCF binding sequences. *Genome Research* **19,** 1338-1349.

Avilion, A. A., Nicolis, S. K., Pevny, L. H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003) Multipotent cell lineages in early mouse development depend on SOX2 function. *Genes Dev.* **17,** 126-140.

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jorgensen, H. F., John, R. M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., and Fisher, A. G. (2006) Chromatin signatures of pluripotent cell lines. *Nat. Cell Biol.* **8,** 532-538.

Bai, S., Ghoshal, K., Datta, J., Majumder, S., Yoon, S. O., and Jacob, S. T. (2005) DNA methyltransferase 3b regulates nerve growth factor-induced differentiation of PC12 cells by recruiting histone deacetylase 2. *Mol. Cell Biol.* **25,** 751-766.

Ballow, D., Meistrich, M. L., Matzuk, M., and Rajkovic, A. (2006) Sohlh1 is essential for spermatogonial differentiation. *Dev. Biol.* **294,** 161-167.

Bannister, A. J., Zegerman, P., Partridge, J. F., Miska, E. A., Thomas, J. O., Allshire, R. C., and Kouzarides, T. (2001) Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120-124.

Bannister, A. J., Schneider, R., and Kouzarides, T. (2002) Histone methylation: dynamic or static? *Cell* **109,** 801-806.

Bao, S. Q., Tang, F. C., Li, X. H., Hayashi, K., Gillich, A., Lao, K. Q., and Surani, M. A. (2009) Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* **461,** 1292-1295.

Barchuk, A. R., Cristino, A. S., Kucharski, R., Costa, L. F., Simoes, Z. L. P., and Maleszka, R. (2007) Molecular determinants of caste differentiation in the highly eusocial honeybee Apis mellifera. *Bmc Dev. Biol.* **7**.

Barr, H., Hermann, A., Berger, J., Tsai, H. H., Adie, K., Prokhortchouk, A., Hendrich, B., and Bird, A. (2007) Mbd2 contributes to DNA methylation-directed repression of the Xist gene. *Mol Cell Biol* **27,** 3750-3757.

Barrera, L. O., Li, Z., Smith, A. D., Arden, K. C., Cavenee, W. K., Zhang, M. Q., Green, R. D., and Ren, B. (2008) Genome-wide mapping and analysis of active promoters in mouse embryonic stem cells and adult organs. *Genome Research* **18,** 46-59.

Barski, A., Cuddapah, S., Cui, K., Roh, T. Y., Schones, D. E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell* **129,** 823-837.

Bartolomei, M. S. (2009) Genomic imprinting: employing and avoiding epigenetic processes. *Genes Dev* **23,** 2124-2133.

Beaujean, N., Taylor, J. E., McGarry, M., Gardner, J. O., Wilmut, I., Loi, P., Ptak, G., Galli, C., Lazzari, G., Bird, A., Young, L. E., and Meehan, R. R. (2004) The effect of interspecific oocytes on demethylation of sperm DNA. *Proc. Natl. Acad. Sci. U.S.A.* **101,** 7636-7640.

Beddington, R. S. and Robertson, E. J. (1989) An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105,** 733-737.

Bell, A. C., West, A. G., and Felsenfeld, G. (1999) The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* **98,** 387-396.

Bell, A. C. and Felsenfeld, G. (2000) Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405,** 482-485.

Berger, J., Sansom, O., Clarke, A., and Bird, A. (2007) MBD2 is required for correct spatial gene expression in the gut. *Mol Cell Biol* **27,** 4049-4057.

Bernstein, B. E., Humphrey, E. L., Erlich, R. L., Schneider, R., Bouman, P., Liu, J. S., Kouzarides, T., and Schreiber, S. L. (2002) Methylation of histone H3 Lys 4 in coding regions of active genes. *Proc. Natl. Acad. Sci. U. S. A* **99,** 8695-8700.

Bernstein, B. E., Mikkelsen, T. S., Xie, X., Kamal, M., Huebert, D. J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., Jaenisch, R., Wagschal, A., Feil, R., Schreiber, S. L., and Lander, E. S. (2006) A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125,** 315-326.

Bernstein, B. E., Meissner, A., and Lander, E. S. (2007) The mammalian epigenome. *Cell* **128,** 669-681.

Bestor, T., Laudano, A., Mattaliano, R., and Ingram, V. (1988) Cloning and sequencing of a cDNA encoding DNA methyltransferase of mouse cells. The carboxyl-terminal domain of the mammalian enzymes is related to bacterial restriction methyltransferases. *J. Mol. Biol.* **203,** 971-983.

Bestor, T. H. and Ingram, V. M. (1985) Growth-dependent expression of multiple species of DNA methyltransferase in murine erythroleukemia cells. *Proc. Natl. Acad. Sci. U. S. A* **82,** 2674-2678.

Bestor, T. H. and Verdine, G. L. (1994) DNA Methyltransferases. *Current Opinion in Cell Biology* **6,** 380-389.

Bestor, T. H. (2000) The DNA methyltransferases of mammals. *Hum. Mol. Gen* **9,** 2395-2402.

Beverdam, A. and Koopman, P. (2006) Expression profiling of purified mouse gonadal somatic cells during the critical time window of sex determination reveals novel candidate genes for human sexual dysgenesis syndromes. *Hum. Mol. Gen* **15,** 417-431.

Biniszkiewicz, D., Gribnau, J., Ramsahoye, B., Gaudet, F., Eggan, K., Humpherys, D., Mastrangelo, M. A., Jun, Z., Walter, J., and Jaenisch, R. (2002) Dnmt1 overexpression causes genomic hypermethylation, loss of imprinting, and embryonic lethality. *Mol. Cell Biol.* **22,** 2124-2135.

Bird, A. and Tweedie, S. (1995) Transcriptional noise and the evolution of gene number. *Philos. Trans. R. Soc. Lond B Biol. Sci.* **349,** 249-253.

Bird, A. (2002) DNA methylation patterns and epigenetic memory. *Genes Dev.* **16,** 6-21.

Bird, A. (2007) Perceptions of epigenetics. *Nature* **447,** 396-398.

Bird, A. P. (1984) DNA Methylation - How Important in Gene-Control. *Nature* **307,** 503-504.

Bird, A. P. (1986) CpG-Rich Islands and the Function of DNA Methylation. *Nature* **321,** 209-213.

Bird, A. P. and Wolffe, A. P. (1999) Methylation-induced repression--belts, braces, and chromatin. *Cell* **99,** 451-454.

Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph,

S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., Andrews, R. M., Flicek, P., Boyle, P. J., Cao, H., Carter, N. P., Clelland, G. K., Davis, S., Day, N., Dhami, P., Dillon, S. C., Dorschner, M. O., Fiegler, H., Giresi, P. G., Goldy, J., Hawrylycz, M., Haydock, A., Humbert, R., James, K. D., Johnson, B. E., Johnson, E. M., Frum, T. T., Rosenzweig, E. R., Karnani, N., Lee, K., Lefebvre, G. C., Navas, P. A., Neri, F., Parker, S. C., Sabo, P. J., Sandstrom, R., Shafer, A., Vetrie, D., Weaver, M., Wilcox, S., Yu, M., Collins, F. S., Dekker, J., Lieb, J. D., Tullius, T. D., Crawford, G. E., Sunyaev, S., Noble, W. S., Dunham, I., Denoeud, F., Reymond, A., Kapranov, P., Rozowsky, J., Zheng, D., Castelo, R., Frankish, A., Harrow, J., Ghosh, S., Sandelin, A., Hofacker, I. L., Baertsch, R., Keefe, D., Dike, S., Cheng, J., Hirsch, H. A., Sekinger, E. A., Lagarde, J., Abril, J. F., Shahab, A., Flamm, C., Fried, C., Hackermuller, J., Hertel, J., Lindemeyer, M., Missal, K., Tanzer, A., Washietl, S., Korbel, J., Emanuelsson, O., Pedersen, J. S., Holroyd, N., Taylor, R., Swarbreck, D., Matthews, N., Dickson, M. C., Thomas, D. J., Weirauch, M. T., Gilbert, J., Drenkow, J., Bell, I., Zhao, X., Srinivasan, K. G., Sung, W. K., Ooi, H. S., Chiu, K. P., Foissac, S., Alioto, T., Brent, M., Pachter, L., Tress, M. L., Valencia, A., Choo, S. W., Choo, C. Y., Ucla, C., Manzano, C., Wyss, C., Cheung, E., Clark, T. G., Brown, J. B., Ganesh, M., Patel, S., Tammana, H., Chrast, J., Henrichsen, C. N., Kai, C., Kawai, J., Nagalakshmi, U., Wu, J., Lian, Z., Lian, J., Newburger, P., Zhang, X., Bickel, P., Mattick, J. S., Carninci, P., Hayashizaki, Y., Weissman, S., Hubbard, T., Myers, R. M., Rogers, J., Stadler, P. F., Lowe, T. M., Wei, C. L., Ruan, Y., Struhl, K., Gerstein, M., Antonarakis, S. E., Fu, Y., Green, E. D., Karaoz, U., Siepel, A., Taylor, J., Liefer, L. A., Wetterstrand, K. A., Good, P. J., Feingold, E. A., Guyer, M. S., Cooper, G. M., Asimenos, G., Dewey, C. N., Hou, M., Nikolaev, S., Montoya-Burgos, J. I., Loytynoja, A., Whelan, S., Pardi, F., Massingham, T., Huang, H., Zhang, N. R., Holmes, I., Mullikin, J. C., Ureta-Vidal, A., Paten, B., Seringhaus, M., Church, D., Rosenbloom, K., Kent, W. J., Stone, E. A., Batzoglou, S., Goldman, N., Hardison, R. C., Haussler, D., Miller, W., Sidow, A., Trinklein, N. D., Zhang, Z. D., Barrera, L., Stuart, R., King, D. C., Ameur, A., Enroth, S., Bieda, M. C., Kim, J., Bhinge, A. A., Jiang, N., Liu, J., Yao, F., Vega, V. B., Lee, C. W., Ng, P., Shahab, A., Yang, A., Moqtaderi, Z., Zhu, Z., Xu, X., Squazzo, S., Oberley, M. J., Inman, D., Singer, M. A., Richmond, T. A., Munn, K. J., Rada-Iglesias, A., Wallerman, O., Komorowski, J., Fowler, J. C., Couttet, P., Bruce, A. W., Dovey, O. M., Ellis, P. D., Langford, C. F., Nix, D. A., Euskirchen, G., Hartman, S., Urban, A. E., Kraus, P., Van Calcar, S., Heintzman, N., Kim, T. H., Wang, K., Qu, C., Hon, G., Luna, R., Glass, C. K., Rosenfeld, M. G., Aldred, S. F., Cooper, S. J., Halees, A., Lin, J. M., Shulha, H. P., Zhang, X., Xu, M., Haidar, J. N., Yu, Y., Ruan, Y., Iyer, V. R., Green, R. D., Wadelius, C., Farnham, P. J., Ren, B., Harte, R. A., Hinrichs, A. S., Trumbower, H., and Clawson, H. (2007) Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447,** 799-816.

Boiani, M., Eckardt, S., Scholer, H. R., and McLaughlin, K. J. (2002) Oct4 distribution and level in mouse clones: consequences for pluripotency. *Genes Dev.* **16,** 1209-1219.

Bostick, M., Kim, J. K., Esteve, P. O., Clark, A., Pradhan, S., and Jacobsen, S. E. (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* **317,** 1760-1764.

Boulikas, T., Wiseman, J. M., and Garrard, W. T. (1980) Points of contact between histone H1 and the histone octamer. *Proc. Natl. Acad. Sci. U. S. A* **77,** 127-131.

Bourc'his, D. and Viegas-Pequignot, E. (2001) Direct analysis of chromosome methylation. *Methods Mol. Biol.* **181,** 229-242.

Bourc'his, D., Xu, G. L., Lin, C. S., Bollman, B., and Bestor, T. H. (2001) Dnmt3L and the establishment of maternal genomic imprints. *Science* **294,** 2536-2539.

Bourc'his, D. and Bestor, T. H. (2004) Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431,** 96-99.

Boyer, L. A., Lee, T. I., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R., and Young, R. A. (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947-956.

Boyer, L. A., Plath, K., Zeitlinger, J., Brambrink, T., Medeiros, L. A., Lee, T. I., Levine, S. S., Wernig, M., Tajonar, A., Ray, M. K., Bell, G. W., Otte, A. P., Vidal, M., Gifford, D. K., Young, R. A., and Jaenisch, R. (2006) Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature* **441,** 349-353.

Boyes, J. and Bird, A. (1991) DNA Methylation Inhibits Transcription Indirectly Via A Methyl-CpG Binding-Protein. *Cell* **64,** 1123-1134.

Bradley, A., Evans, M., Kaufman, M. H., and Robertson, E. (1984) Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. *Nature* **309,** 255-256.

Brandeis, M., Kafri, T., Ariel, M., Chaillet, J. R., Mccarrey, J., Razin, A., and Cedar, H. (1993) The Ontogeny of Allele-Specific Methylation Associated with Imprinted Genes in the Mouse. *EMBO Journal* **12,** 3669-3677.

Brenner, C. and Fuks, F. (2006) DNA methyltransferases: facts, clues, mysteries. *Curr. Top. Microbiol. Immunol.* **301,** 45-66.

Brons, I. G. M., Smithers, L. E., Trotter, M. W. B., Rugg-Gunn, P., Sun, B. W., Lopes, S. M. C. D., Howlett, S. K., Clarkson, A., Ahrlund-Richter, L., Pedersen, R. A., and Vallier, L. (2007) Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448,** 191-1U7.

Brown, S. W. (1966) Heterochromatin. *Science* **151,** 417-425.

Bruce, K., Myers, F. A., Mantouvalou, E., Lefevre, P., Greaves, I., Bonifer, C., Tremethick, D. J., Thorne, A. W., and Crane-Robinson, C. (2005) The replacement histone H2A.Z in a hyperacetylated form is a feature of active genes in the chicken. *Nucleic Acids Res.* **33,** 5633-5639.

Buschbeck, M., Uribesalgo, I., Wibowo, I., Rue, P., Martin, D., Gutierrez, A., Morey, L., Guigo, R., Lopez-Schier, H., and Di Croce, L. (2009) The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nat. Struct. Mol. Biol.* **16,** 1074-1079.

Caballero, O. L. and Chen, Y. T. (2009) Cancer/testis (CT) antigens: Potential targets for immunotherapy. *Cancer Science* **100,** 2014-2021.

Campanero, M. R., Armstrong, M. I., and Flemington, E. K. (2000) CPG methylation as a mechanism for the regulation of E2F activity. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 6481-6486.

Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., Jones, R. S., and Zhang, Y. (2002) Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298,** 1039-1043.

Cedar, H. and Bergman, Y. (2009) Linking DNA methylation and histone modification: patterns and paradigms. *Nat. Rev. Genet.* **10,** 295-304.

Chadwick, B. P. and Willard, H. F. (2001) A novel chromatin protein, distantly related to histone H2A, is largely excluded from the inactive X chromosome. *J. Cell Biol.* **152,** 375-384.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003) Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113,** 643-655.

Chambers, I. and Smith, A. (2004) Self-renewal of teratocarcinoma and embryonic stem cells. *Oncogene* **23,** 7150-7160.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007) Nanog safeguards pluripotency and mediates germline development. *Nature* **450,** 1230-12U8.

Chan, M. F., van Amerongen, R., Nijjar, T., Cuppen, E., Jones, P. A., and Laird, P. W. (2001) Reduced rates of gene loss, gene silencing, and gene mutation in Dnmt1-deficient embryonic stem cells. *Mol. Cell Biol.* **21,** 7587-7600.

Chandler, V. L. (2007) Paramutation: from maize to mice. *Cell* **128,** 641-645.

Chaturvedi, C. P., Hosey, A. M., Palii, C., Perez-Iratxeta, C., Nakatani, Y., Ranish, J. A., Dilworth, F. J., and Brand, M. (2009) Dual role for the methyltransferase G9a in the maintenance of beta-globin gene transcription in adult erythroid cells. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 18303-18308.

Chen, D., Ma, H., Hong, H., Koh, S. S., Huang, S. M., Schurter, B. T., Aswad, D. W., and Stallcup, M. R. (1999) Regulation of transcription by a protein methyltransferase. *Science* **284,** 2174-2177.

Chen, R. Z., Pettersson, U., Beard, C., Jackson-Grusby, L., and Jaenisch, R. (1998) DNA hypomethylation leads to elevated mutation rates. *Nature* **395,** 89-93.

Chen, R. Z., Akbarian, S., Tudor, M., and Jaenisch, R. (2001) Deficiency of methyl-CpG binding protein-2 in CNS neurons results in a Rett-like phenotype in mice. *Nature Genetics* **27,** 327-331.

Chen, T., Ueda, Y., Dodge, J. E., Wang, Z., and Li, E. (2003) Establishment and maintenance of genomic methylation patterns in mouse embryonic stem cells by Dnmt3a and Dnmt3b. *Mol. Cell Biol.* **23,** 5594-5605.

Chen, T. and Li, E. (2006) Establishment and maintenance of DNA methylation patterns in mammals. *Curr. Top. Microbiol. Immunol.* **301,** 179-201.

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y., and Li, E. (2007) Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nat. Genet.* **39,** 391-396.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V. B., Wong, E., Orlov, Y. L., Zhang, W., Jiang, J., Loh, Y. H., Yeo, H. C., Yeo, Z. X., Narang, V., Govindarajan, K. R., Leong, B., Shahab, A., Ruan, Y., Bourque, G., Sung, W. K., Clarke, N. D., Wei, C. L., and Ng, H. H. (2008) Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133,** 1106-1117.

Chen, Z. F., Paquette, A. J., and Anderson, D. J. (1998) NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nature Genetics* **20,** 136-142.

Chew, J. L., Loh, Y. H., Zhang, W. S., Chen, X., Tam, W. L., Yeap, L. S., Li, P., Ang, Y. S., Lim, B., Robson, P., and Ng, H. H. (2005) Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. *Mol Cell Biol* **25,** 6031-6046.

Chotalia, M., Smallwood, S. A., Ruf, N., Dawson, C., Lucifero, D., Frontera, M., James, K., Dean, W., and Kelsey, G. (2009) Transcription is required for establishment of germline methylation marks at imprinted genes. *Genes Dev* **23,** 105-117.

Chow, C. M., Georgiou, A., Szutorisz, H., Maia e Silva, Pombo, A., Barahona, I., Dargelos, E., Canzonetta, C., and Dillon, N. (2005) Variant histone H3.3 marks promoters of transcriptionally active genes during mammalian cell division. *EMBO Rep.* **6,** 354-360.

Chuang, L. S., Ian, H. I., Koh, T. W., Ng, H. H., Xu, G., and Li, B. F. (1997) Human DNA-(cytosine-5) methyltransferase-PCNA complex as a target for p21WAF1. *Science* **277,** 1996-2000.

Ciccone, D. N., Su, H., Hevi, S., Gay, F., Lei, H., Bajko, J., Xu, G. L., Li, E., and Chen, T. P. (2009) KDM1B is a histone H3K4 demethylase required to establish maternal genomic imprints. *Nature* **461,** 415-U115.

Clarkson, M. J., Wells, J. R. E., Gibson, F., Saint, R., and Tremethick, D. J. (1999) Regions of variant histone His2AvD required for Drosophila development. *Nature* **399,** 694-697.

Coe, E. H., Jr. (1968) Heritable repression due to paramutation in maize. *Science* **162:** 925

Collins, T., Stone, J. R., and Williams, A. J. (2001) All in the family: the BTB/POZ, KRAB, and SCAN domains. *Mol Cell Biol* **21,** 3609-3615.

Costanzi, C. and Pehrson, J. R. (1998) Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals. *Nature* **393,** 599-601.
Craig, J. M. and Bickmore, W. A. (1994) The distribution of CpG islands in mammalian chromosomes. *Nat. Genet.* **7,** 376-382.

Cross, S. H., Meehan, R. R., Nan, X. S., and Bird, A. (1997) A component of the transcriptional repressor MeCP1 shares a motif with DNA methyltransferase and HRX proteins. *Nature Genetics* **16,** 256-259.

Csankovszki, G., Nagy, A., and Jaenisch, R. (2001) Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J. Cell Biol.* **153,** 773-784.

Czermin, B., Melfi, R., McCabe, D., Seitz, V., Imhof, A., and Pirrotta, V. (2002) Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111,** 185-196.

Damelin, M. and Bestor, T. H. (2007) Biological functions of DNA methyltransferase 1 require its methyltransferase activity. *Mol. Cell Biol.* **27,** 3891-3899.

Daniel, J. M. and Reynolds, A. B. (1999) The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. *Mol Cell Biol* **19,** 3614-3623.

De La, F. R., Baumann, C., Fan, T., Schmidtmann, A., Dobrinski, I., and Muegge, K. (2006) Lsh is required for meiotic chromosome synapsis and retrotransposon silencing in female germ cells. *Nat. Cell Biol.* **8,** 1448-1454.

de Napoles, M., Nesterova, T., and Brockdorff, N. (2007) Early loss of Xist RNA expression and inactive X chromosome associated chromatin modification in developing primordial germ cells. *PLoS One.* **2,** e860.

De Smet, C., De Backer, O., Faraoni, I., Lurquin, C., Brasseur, F., and Boon, T. (1996) The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. *Proc. Natl. Acad. Sci. U. S A* **93,** 7149-7153.

De Smet, C., Lurquin, C., Lethe, B., Martelange, V., and Boon, T. (1999) DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol Cell Biol* **19,** 7327-7335.

Dean, W., Santos, F., Stojkovic, M., Zakhartchenko, V., Walter, J., Wolf, E., and Reik, W. (2001) Conservation of methylation reprogramming in mammalian development: aberrant reprogramming in cloned embryos. *Proc. Natl. Acad. Sci. U. S. A* **98,** 13734-13738.

Deb-Rinker, P., Ly, D., Jezierski, A., Sikorska, M., and Walker, P. R. (2005) Sequential DNA methylation of the Nanog and Oct-4 upstream regions in human NT2 cells during neuronal differentiation. *J. Biol. Chem.* **280,** 6257-6260.

Defossez, P. A., Kelly, K. F., Filion, G. J. P., Perez-Torrado, R., Magdinier, F., Menoni, H., Nordgaard, C. L., Daniel, J. M., and Gilson, E. (2005) The human enhancer blocker CTC-binding factor interacts with the transcription factor Kaiso. *J Biol Chem.* **280,** 43017-43023.

Della Ragione, F., Tiunova, A., Vacca, M., Strazzullo, M., Gonzalez, E., Armstrong, J., Valero, R., Campanile, C., Pineda, M., Hulten, M., Monros, E., D'Esposito, M., and Prokhortchouk, E. (2006) The X-linked methyl binding protein gene Kaiso is highly expressed in brain but is not mutated in Rett syndrome patients. *Gene* **373,** 83-89.

Dennis, K., Fan, T., Geiman, T., Yan, Q. S., and Muegge, K. (2001) Lsh, a member of the SNF2 family, is required for genome-wide methylation. *Genes Dev* **15,** 2940-2944.

Deplus, R., Brenner, C., Burgers, W. A., Putmans, P., Kouzarides, T., de Launoit, Y., and Fuks, F. (2002) Dnmt3L is a transcriptional repressor that recruits histone deacetylase. *Nucleic Acids Res.* **30,** 3831-3838.

Dhalluin, C., Carlson, J. E., Zeng, L., He, C., Aggarwal, A. K., and Zhou, M. M. (1999) Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399,** 491-496.

Ding, F. and Chaillet, J. R. (2002) In vivo stabilization of the Dnmt1 (cytosine-5)-methyltransferase protein. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 14861-14866.

Dodge, J. E., Okano, M., Dick, F., Tsujimoto, N., Chen, T. P., Wang, S. M., Ueda, Y., Dyson, N., and Li, E. (2005) Inactivation of Dnmt3b in mouse embryonic fibroblasts results in DNA hypomethylation, chromosomal instability, and spontaneous immortalization. *J Biol Chem.* **280,** 17986-17991.

Doherty, A. S., Bartolomei, M. S., and Schultz, R. M. (2002) Regulation of stage-specific nuclear translocation of Dnmt1o during preimplantation mouse development. *Dev. Biol.* . **242,** 255-266.

Donaldson, A. D. (2005) Shaping time: chromatin structure and the DNA replication programme. *Trends Genet.* **21,** 444-449.

Dong, K. B., Maksakova, I. A., Mohn, F., Leung, D., Appanah, R., Lee, S., Yang, H. W., Lam, L. L., Mager, D. L., Schubeler, D., Tachibana, M., Shinkai, Y., and Lorincz, M. C. (2008) DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO Journal* **27,** 2691-2701.

Downing, G. J. and Battey, J. F., Jr. (2004) Technical assessment of the first 20 years of research using mouse embryonic stem cell lines. *Stem Cells* **22,** 1168-1180.

Drosatos, K., Sanoudou, D., Kypreos, K. E., Kardassis, D., and Zannis, V. I. (2007) A dominant negative form of the transcription factor c-Jun affects genes that have opposing effects on lipid homeostasis in mice. *J Biol Chem.* **282,** 19556-19564.

Drummond, J. T. and Bellacosa, A. (2001) Human DNA mismatch repair in vitro operates independently of methylation status at CpG sites. *Nucleic Acids Res.* **29,** 2234-2243.

Dunican, D. S., Ruzov, A., Hackett, J. A., and Meehan, R. R. (2008) xDnmt1 regulates transcriptional silencing in pre-MBT Xenopus embryos independently of its catalytic function. *Development* **135,** 1295-1302.

Durcova-Hills, G., Ainscough, J., and Mclaren, A. (2001) Pluripotential stem cells derived from migrating primordial germ cells. *Differentiation* **68,** 220-226.

Eads, C. A., Danenberg, K. D., Kawakami, K., Saltz, L. B., Danenberg, P. V., and Laird, P. W. (1999) CpG island hypermethylation in human colorectal tumors is not associated with DNA methyltransferase overexpression. *Cancer Res.* **59,** 2302-2306.

Easwaran, H. P., Schermelleh, L., Leonhardt, H., and Cardoso, M. C. (2004) Replication-independent chromatin loading of Dnmt1 during G2 and M phases. *EMBO Rep.* **5,** 1181-1186.

Eckhardt, F., Lewin, J., Cortese, R., Rakyan, V. K., Attwood, J., Burger, M., Burton, J., Cox, T. V., Davies, R., Down, T. A., Haefliger, C., Horton, R., Howe, K., Jackson, D. K., Kunde, J., Koenig, C., Liddle, J., Niblett, D., Otto, T., Pettett, R., Seemann, S., Thompson, C., West, T., Rogers, J., Olek, A., Berlin, K., and Beck, S. (2006) DNA methylation profiling of human chromosomes 6, 20 and 22. *Nature Genetics* **38,** 1378-1385.

Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003) Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300,** 455.

Eggan, E., Baldwin, K., Tackett, M., Osborne, J., Gogos, J., Chess, A., Axel, R., and Jaenisch, R. (2004) Mice cloned from olfactory sensory neurons. *Nature* **428,** 44-49.

Eggan, K., Akutsu, H., Loring, J., Jackson-Grusby, L., Klemm, M., Rideout, W. M., III, Yanagimachi, R., and Jaenisch, R. (2001) Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl. Acad. Sci. U. S. A* **98,** 6209-6214.

Egger, G., Liang, G. N., Aparicio, A., and Jones, P. A. (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature* **429,** 457-463.

Egger, G., Jeong, S., Escobar, S. G., Cortez, C. C., Li, T. W. H., Saito, Y., Yoo, C. B., Jones, P. A., and Liang, G. N. (2006) Identification of DNMT1 (DNA methyltransferase 1) hypomorphs in somatic knockouts suggests an essential role for DNMT1 in cell survival. *Proc. Natl. Acad. Sci. U.S.A.* **103,** 14080-14085.

Ehrlich, M. (2003) Expression of various genes is controlled by DNA methylation during mammalian development. *J. Cell Biochem.* **88,** 899-910.

el Deiry, W. S., Nelkin, B. D., Celano, P., Yen, R. W., Falco, J. P., Hamilton, S. R., and Baylin, S. B. (1991) High expression of the DNA methyltransferase gene characterizes human neoplastic cells and progression stages of colon cancer. *Proc. Natl. Acad. Sci. U. S. A* **88,** 3470-3474.

El Osta, A. (2003) DNMT cooperativity--the developing links between methylation, chromatin structure and cancer. *Bioessays* **25,** 1071-1084.

Eminli, S., Utikal, J., Arnold, K., Jaenisch, R., and Hochedlinger, K. (2008) Reprogramming of Neural Progenitor Cells into Induced Pluripotent Stem Cells in the Absence of Exogenous Sox2 Expression. *Stem Cells* **26,** 2467-2474.

Epsztejn-Litman, S., Feldman, N., Abu-Remaileh, M., Shufaro, Y., Gerson, A., Ueda, J., Deplus, R., Fuks, F., Shinkai, Y., Cedar, H., and Bergman, Y. (2008) De novo DNA methylation promoted by G9a prevents reprogramming of embryonically silenced genes. *Nature Structural & Molecular Biology* **15,** 1176-1183.

Esteller, M. (2005) Dormant hypermethylated tumour suppressor genes: questions and answers. *J. Pathol.* **205,** 172-180.

Esteve, P. O., Patnaik, D., Chin, H. G., Benner, J., Teitell, M. A., and Pradhan, S. (2005) Functional analysis of the N- and C-terminus of mammalian G9a histone H3 methyltransferase. *Nucleic Acids Res.* **33,** 3211-3223.

Esteve, P. O., Chin, H. G., and Pradhan, S. (2005) Human maintenance DNA (cytosine-5)-methyltransferase and p53 modulate expression of p53-repressed promoters. *Proc. Natl. Acad. Sci. U. S. A* **102,** 1000-1005.

Evans, M. J. (1972) The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. *J. Embryol. Exp. Morphol.* **28,** 163-176.

Evans, M. J. and Kaufman, M. H. (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292,** 154-156.

Faast, R., Thonglairoam, V., Schulz, T. C., Beall, J., Wells, J. R. E., Taylor, H., Matthaei, K., Rathjen, P. D., Tremethick, D. J., and Lyons, I. (2001) Histone variant H2A.Z is required for early mammalian development. *Current Biology* **11,** 1183-1187.

Fan, H. Y., Trotter, K. W., Archer, T. K., and Kingston, R. E. (2005) Swapping function of two chromatin remodeling complexes. *Mol. Cell* **17,** 805-815.

Fan, T., Hagan, J. P., Kozlov, S. V., Stewart, C. L., and Muegge, K. (2005) Lsh controls silencing of the imprinted Cdkn1c gene. *Development* **132,** 635-644.

Fang, J. Y., Lu, J., Chen, Y. X., and Yang, L. (2003) Effects of DNA methylation on expression of tumor suppressor genes and proto-oncogene in human colon cancer cell lines. *World J. Gastroenterol.* **9,** 1976-1980.

Farris, S. D., Rubio, E. D., Moon, J. J., Gombert, W. M., Nelson, B. H., and Krumm, A. (2005) Transcription-induced chromatin remodeling at the c-myc gene involves the local exchange of histone H2A.Z. *J Biol Chem.* **280,** 25298-25303.

Farthing, C. R., Ficz, G., Ng, R. K., Chan, C. F., Andrews, S., Dean, W., Hemberger, M., and Reik, W. (2008) Global Mapping of DNA Methylation in Mouse Promoters Reveals Epigenetic Reprogramming of Pluripotency Genes. *Plos Genetics* **4**.

Fatemi, M., Hermann, A., Pradhan, S., and Jeltsch, A. (2001) The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *Journal of Molecular Biology* **309,** 1189-1199.

Feldman, N., Gerson, A., Fang, J., Li, E., Zhang, Y., Shinkai, Y., Cedar, H., and Bergman, Y. (2006) G9a-mediated irreversible epigenetic inactivation of Oct-3/4 during early embryogenesis. *Nat. Cell Biol.* **8,** 188-194.

Felsenfeld, G. and Mcghee, J. (1982) Methylation and Gene-Control. *Nature* **296,** 602-603.

Feng, Y. Q., Desprat, R., Fu, H., Olivier, E., Lin, C. M., Lobell, A., Gowda, S. N., Aladjem, M. I., and Bouhassira, E. E. (2006) DNA methylation supports intrinsic epigenetic memory in mammalian cells. *Plos Genetics* **2,** 461-470.

Feng, B., Jiang, J. M., Kraus, P., Ng, J. H., Heng, J. C. D., Chan, Y. S., Yaw, L. P., Zhang, W. W., Loh, Y. H., Han, J. Y., Vega, V. B., Cacheux-Rataboul, V., Lim, B., Lufkin, T., and Ng, H. H. (2009) Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. *Nature Cell Biology* **11,** 197-U193.

Field, L. M., Lyko, F., Mandrioli, M., and Prantera, G. (2004) DNA methylation in insects. *Insect Molecular Biology* **13,** 109-115.

Filion, G. J. P., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E., and Defossez, P. A. (2006) A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol Cell Biol* **26,** 169-181.

Finch, J. T. and Klug, A. (1976) Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci. U. S. A* **73,** 1897-1901.

Flanagan, J. F., Mi, L. Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005) Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* **438,** 1181-1185.

Fouse, S. D., Shen, Y., Pellegrini, M., Cole, S., Meissner, A., Van Neste, L., Jaenisch, R., and Fan, G. P. (2008) Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3K4/K27 trimethylation. *Cell Stem Cell* **2,** 160-169.

Fraga, M. F., Ballestar, E., Montoya, G., Taysavang, P., Wade, P. A., and Esteller, M. (2003) The affinity of different MBD proteins for a specific methylated locus depends on their intrinsic binding properties. *Nucleic Acids Res.* **31,** 1765-1774.

Fraser, P. and Bickmore, W. (2007) Nuclear organization of the genome and the potential for gene regulation. *Nature* **447,** 413-417.

Fuhrmann, G., Chung, A. C. K., Jackson, K. J., Hummelke, G., Baniahmad, A., Sutter, J., Sylvester, I., Scholer, H. R., and Cooney, A. J. (2001) Mouse germline restriction of Oct4 expression by germ cell nuclear factor. *Developmental Cell* **1,** 377-387.

Fujita, N., Watanabe, S., Ichimura, T., Tsuruzoe, S., Shinkai, Y., Tachibana, M., Chiba, T., and Nakao, M. (2003) Methyl-CpG binding domain 1 (MBD1) interacts with the Suv39h1-HP1 heterochromatic complex for DNA methylation-based transcriptional repression. *J Biol Chem.* **278,** 24132-24138.

Fukazawa, H., Nakano, S., Mizuno, S., and Uehara, Y. (1996) Inhibitors of anchorage-independent growth affect the growth of transformed cells on poly(2-hydroxyethyl methacrylate)-coated surfaces. *Int. J. Cancer* **67,** 876-882.

Fuks, F., Burgers, W. A., Brehm, A., Hughes-Davies, L., and Kouzarides, T. (2000) DNA methyltransferase Dnmt1 associates with histone deacetylase activity. *Nat. Genet.* **24,** 88-91.

Fuks, F. (2003) [DNA methyltransferases: from chromatin remodeling to cancer]. *Med. Sci. (Paris)* **19,** 477-480.

Fuks, F., Hurd, P. J., Deplus, R., and Kouzarides, T. (2003) The DNA methyltransferases associate with HP1 and the SUV39H1 histone methyltransferase. *Nucleic Acids Res.* **31,** 2305-2312.

Futscher, B. W., Oshiro, M. M., Wozniak, R. J., Holtan, N., Hanigan, C. L., Duan, H., and Domann, F. E. (2002) Role for DNA methylation in the control of cell type specific maspin expression. *Nat. Genet.* **31,** 175-179.

Gagnon-Kugler, T., Langlois, F., Stefanovsky, V., Lessard, F., and Moss, T. (2009) Loss of human ribosomal gene CpG methylation enhances cryptic RNA polymerase II transcription and disrupts ribosomal RNA processing. *Mol. Cell* **35,** 414-425.

Gaudet, F., Hodgson, J. G., Eden, A., Jackson-Grusby, L., Dausman, J., Gray, J. W., Leonhardt, H., and Jaenisch, R. (2003) Induction of tumors in mice by genomic hypomethylation. *Science* **300,** 489-492.

Gaudet, F., Rideout, W. M., III, Meissner, A., Dausman, J., Leonhardt, H., and Jaenisch, R. (2004) Dnmt1 expression in pre- and postimplantation embryogenesis and the maintenance of IAP silencing. *Mol. Cell Biol.* **24,** 1640-1648.

Geijsen, N., Horoschak, M., Kim, K., Gribnau, J., Eggan, K., and Daley, G. Q. (2004) Derivation of embryonic germ cells and male gametes from embryonic stem cells. *Nature* **427,** 148-154.

Geiman, T. M., Tessarollo, L., Anver, M. R., Kopp, J. B., Ward, J. M., and Muegge, K. (2001) Lsh, a SNF2 family member, is required for normal murine development. *Biochimica et Biophysica Acta-General Subjects* **1526,** 211-220.

Geiman, T. M., Sankpal, U. T., Robertson, A. K., Chen, Y., Mazumdar, M., Heale, J. T., Schmiesing, J. A., Kim, W., Yokomori, K., Zhao, Y., and Robertson, K. D. (2004) Isolation and characterization of a novel DNA methyltransferase complex linking DNMT3B with components of the mitotic chromosome condensation machinery. *Nucleic Acids Res.* **32,** 2716-2729.

Ghoshal, K., Datta, J., Majumder, S., Bai, S. M., Kutay, H., Motiwala, T., and Jacob, S. T. (2005) 5-Aza-deoxycytidine induces selective degradation of DNA methyltransferase 1 by a proteasomal pathway that requires the KEN box, bromo-adjacent homology domain, and nuclear localization signal. *Mol Cell Biol* **25,** 4727-4741.

Gilbert, N., Thomson, I., Boyle, S., Allan, J., Ramsahoye, B., and Bickmore, W. A. (2007) DNA methylation affects nuclear organization, histone modifications, and linker histone binding but not chromatin compaction. *J. Cell Biol.* **177,** 401-411.

Ginsburg, M., Snow, M. H. L., and Mclaren, A. (1990) Primordial Germ-Cells in the Mouse Embryo During Gastrulation. *Development* **110,** 521-&.

Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007) Epigenetics: a landscape takes shape. *Cell* **128,** 635-638.

Goll, M. G. and Bestor, T. H. (2005) Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74,** 481-514.

Goll, M. G., Kirpekar, F., Maggert, K. A., Yoder, J. A., Hsieh, C. L., Zhang, X., Golic, K. G., Jacobsen, S. E., and Bestor, T. H. (2006) Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2. *Science* **311,** 395-398.

Govin, J., Escoffier, E., Rousseaux, S., Kuhn, L., Ferro, M., Thevenon, J., Catena, R., Davidson, I., Garin, J., Khochbin, S., and Caron, C. (2007) Pericentric heterochromatin reprogramming by new histone variants during mouse spermiogenesis. *Journal of Cell Biology* **176,** 283-294.

Gowher, H., Stockdale, C. J., Goyal, R., Ferreira, H., Owen-Hughes, T., and Jeltsch, A. (2005) De novo methylation of nucleosomal DNA by the mammalian Dnmt1 and Dnmt3A DNA methyltransferases. *Biochemistry* **44,** 9899-9904.

Grange, T., Cappabianca, L., Flavin, M., Sassi, H., and Thomassin, H. (2001) In vivo analysis of the model tyrosine aminotransferase gene reveals multiple sequential steps in glucocorticoid receptor action. *Oncogene* **20,** 3028-3038.

Green, C. M. and Almouzni, G. (2003) Local action of the chromatin assembly factor CAF-1 at sites of nucleotide excision repair in vivo. *EMBO J.* **22,** 5163-5174.

Gregory, T. R. (2001) Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biol. Rev. Camb. Philos. Soc.* **76,** 65-101.

Gregory, T. R. (2002) Genome size and developmental complexity. *Genetica* **115,** 131-146.

Gruenbaum, Y., Stein, R., Cedar, H., and Razin, A. (1981) Methylation of CpG Sequences in Eukaryotic DNA. *Febs Letters* **124,** 67-71.

Gu, P., LeMenuet, D., Chung, A. C., Mancini, M., Wheeler, D. A., and Cooney, A. J. (2005) Orphan nuclear receptor GCNF is required for the repression of pluripotency genes during retinoic acid-induced embryonic stem cell differentiation. *Mol. Cell Biol.* **25,** 8507-8519.

Guenther, M. G., Levine, S. S., Boyer, L. A., Jaenisch, R., and Young, R. A. (2007) A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130,** 77-88.

Guo, G., Wang, W., and Bradley, A. (2004) Mismatch repair genes identified using genetic screens in Blm-deficient embryonic stem cells. *Nature* **429,** 891-895.

Gurdon, J. B. (1962) Adult frogs derived from the nuclei of single somatic cells. *Dev. Biol.* . **4,** 256-273.

Gurdon, J. B. (1992) The generation of diversity and pattern in animal development. *Cell* **68,** 185-199.

Guy, J., Hendrich, B., Holmes, M., Martin, J. E., and Bird, A. (2001) A mouse Mecp2-null mutation causes neurological symptoms that mimic Rett syndrome. *Nat. Genet.* **27,** 322-326.

Hajkova, P., Erhardt, S., Lane, N., Haaf, T., El Maarri, O., Reik, W., Walter, J., and Surani, M. A. (2002) Epigenetic reprogramming in mouse primordial germ cells. *Mechanisms of Development* **117,** 15-23.

Hajkova, P., Ancelin, K., Waldmann, T., Lacoste, N., Lange, U. C., Cesari, F., Lee, C., Almouzni, G., Schneider, R., and Surani, M. A. (2008) Chromatin dynamics during epigenetic reprogramming in the mouse germ line. *Nature* **452,** 877-8U6.

Hake, S. B., Garcia, B. A., Duncan, E. M., Kauer, M., Dellaire, G., Shabanowitz, J., Bazett-Jones, D. P., Allis, C. D., and Hunt, D. F. (2006) Expression patterns and post-translational modifications associated with mammalian histone H3 variants. *J Biol Chem.* **281,** 559-568.

Hammoud, S. S., Nix, D. A., Zhang, H. Y., Purwar, J., Carrell, D. T., and Cairns, B. R. (2009) Distinctive chromatin in human sperm packages genes for embryo development. *Nature* **460,** 473-U47.

Hanna, J., Wernig, M., Markoulaki, S., Sun, C. W., Meissner, A., Cassady, J. P., Beard, C., Brambrink, T., Wu, L. C., Townes, T. M., and Jaenisch, R. (2007) Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin. *Science* **318,** 1920-1923.

Hanna, J., Markoulaki, S., Schorderet, P., Carey, B. W., Beard, C., Wernig, M., Creyghton, M. P., Steine, E. J., Cassady, J. P., Foreman, R., Lengner, C. J., Dausman, J. A., and Jaenisch, R. (2008) Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133,** 250-264.

Hare, J. T. and Taylor, J. H. (1985) One role for DNA methylation in vertebrate cells is strand discrimination in mismatch repair. *Proc. Natl. Acad. Sci. U. S. A* **82,** 7350-7354.

Harfe, B. D. and Jinks-Robertson, S. (2000) DNA mismatch repair and genetic instability. *Annu. Rev. Genet.* **34,** 359-399.

Hark, A. T., Schoenherr, C. J., Katz, D. J., Ingram, R. S., Levorse, J. M., and Tilghman, S. M. (2000) CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405,** 486-489.

Hata, K., Okano, M., Lei, H., and Li, E. (2002) Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development* **129,** 1983-1993.

Hattori, N., Imao, Y., Nishino, K., Hattori, N., Ohgane, J., Yagi, S., Tanaka, S., and Shiota, K. (2007) Epigenetic regulation of Nanog gene in embryonic stem and trophoblast stem cells. *Genes to Cells* **12,** 387-396.

Hayashi, K., Lopes, S. M. C. D., Tang, F., and Surani, M. A. (2008) Dynamic Equilibrium and Heterogeneity of Mouse Pluripotent Stem Cells with Distinct Functional and Epigenetic States. *Cell Stem Cell* **3,** 391-401.

Hayashi, S. and McMahon, A. P. (2002) Efficient recombination in diverse tissues by a tamoxifen-inducible form of Cre: a tool for temporally regulated gene activation/inactivation in the mouse. *Dev. Biol.* . **244,** 305-318.

Heard, E., Clerc, P., and Avner, P. (1997) X-chromosome inactivation in mammals. *Annu. Rev. Genet.* **31,** 571-610.

Hebbes, T. R., Thorne, A. W., and Crane-Robinson, C. (1988) A direct link between core histone acetylation and transcriptionally active chromatin. *EMBO J.* **7,** 1395-1402.

Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A., Wang, W., Weng, Z., Green, R. D., Crawford, G. E., and Ren, B. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39,** 311-318.

Hemberger, M., Dean, W., and Reik, W. (2009) Epigenetic dynamics of stem cells and cell lineage commitment: digging Waddington's canal. *Nat. Rev. Mol. Cell Biol.* **10,** 526-537.

Hendrich, B. and Bird, A. (1998) Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18,** 6538-6547.

Hendrich, B., Guy, J., Ramsahoye, B., Wilson, V. A., and Bird, A. (2001) Closely related proteins MBD2 and MBD3 play distinctive but interacting roles in mouse development. *Genes Dev.* **15,** 710-723.

Hermann, A., Schmitt, S., and Jeltsch, A. (2003) The human Dnmt2 has residual DNA-(cytosine-C5) methyltransferase activity. *J. Biol. Chem.* **278,** 31717-31721.

Hermann, A., Gowher, H., and Jeltsch, A. (2004) Biochemistry and biology of mammalian DNA methyltransferases. *Cell Mol. Life Sci.* **61,** 2571-2587.

Hochedlinger, K., Yamada, Y., Beard, C., and Jaenisch, R. (2005) Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121,** 465-477.

Hochedlinger, K. and Jaenisch, R. (2006) Nuclear reprogramming and pluripotency. *Nature* **441,** 1061-1067.

Hochedlinger, K. and Plath, K. (2009) Epigenetic reprogramming and induced pluripotency. *Development* **136,** 509-523.

Hogan B, Beddington R, Costantini F, Lacy E. (1994). Manipulating the mouse embryo: A laboratory manual. 2nd edition. Cold Spring Harbor, NY: *Cold Spring Harbor Laboratory Press*
Holliday, R. and Pugh, J. E. (1975) DNA modification mechanisms and gene activity during development. *Science* **187,** 226-232.

Holliday, R. (1987) The Inheritance of Epigenetic Defects. *Science* **238,** 163-170.

Holliday, R. (1996) DNA methylation in eukaryotes: 20 years on. In *Epigenetic mechanisms of gene regulation* (ed. W. E. A Russo, R. A. Martienseen & A. D. Riggs). Pp 5-27.

Hotchkiss, R.D. (1948) The quantitative separation of purines, pyrimidines,nucleosides by paper chromatography. *J Biol Chem*, **175**, 315-332.

Howell, C. Y., Bestor, T. H., Ding, F., Latham, K. E., Mertineit, C., Trasler, J. M., and Chaillet, J. R. (2001) Genomic imprinting disrupted by a maternal effect mutation in the Dnmt1 gene. *Cell* **104,** 829-838.

Huang, S., Litt, M., and Felsenfeld, G. (2005) Methylation of histone H4 by arginine methyltransferase PRMT1 is essential in vivo for many subsequent histone modifications. *Genes Dev.* **19,** 1885-1893.

Hublitz, P., Albert, M., and Peters, A. H. (2009) Mechanisms of transcriptional repression by histone lysine methylation. *Int. J. Dev. Biol.* . **53,** 335-354.

Hutchins, A. S., Mullen, A. C., Lee, H. W., Sykes, K. J., High, F. A., Hendrich, B. D., Bird, A. P., and Reiner, S. L. (2002) Gene silencing quantitatively controls the function of a developmental trans-activator. *Mol. Cell* **10,** 81-91.

Iannello, R. C., Gould, J. A., Young, J. C., Giudice, A., Medcalf, R., and Kola, I. (2000) Methylation-dependent silencing of the testis-specific Pdha-2 basal promoter occurs through selective targeting of an activating transcription factor/cAMP-responsive element-binding site. *J Biol Chem.* **275,** 19603-19608.

Iguchiariga, S. M. M. and Schaffner, W. (1989) CpG Methylation of the Camp-Responsive Enhancer Promoter Sequence Tgacgtca Abolishes Specific Factor Binding As Well As Transcriptional Activation. *Genes Dev* **3,** 612-619.

Iida, T., Suetake, I., Tajima, S., Morioka, H., Ohta, S., Obuse, C., and Tsurimoto, T. (2002) PCNA clamp facilitates action of DNA cytosine methyltransferase 1 on hemimethylated DNA. *Genes Cells* **7,** 997-1007.

Ikegami, K., Iwatani, M., Suzuki, M., Tachibana, M., Shinkai, Y., Tanaka, S., Greally, J. M., Yagi, S., Hattori, N., and Shiota, K. (2007) Genome-wide and locus-specific DNA hypomethylation in G9a deficient mouse embryonic stem cells. *Genes Cells* **12,** 1-11.

Illingworth, R., Kerr, A., DeSousa, D., Jorgensen, H., Ellis, P., Stalker, J., Jackson, D., Clee, C., Plumb, R., Rogers, J., Humphray, S., Cox, T., Langford, C., and Bird, A. (2008) A novel CpG island set identifies tissue-specific methylation at developmental gene loci. *Plos Biology* **6,** 37-51.

Illingworth, R. S. and Bird, A. P. (2009) CpG islands - 'A rough guide'. *Febs Letters* **583,** 1713-1720.

Imamura, M., Miura, K., Iwabuchi, K., Ichisaka, T., Nakagawa, M., Lee, J., Kanatsu-Shinohara, M., Shinohara, T., and Yamanaka, S. (2006) Transcriptional repression and DNA hypermethylation of a small set of ES cell marker genes in male germline stem cells. *Bmc Dev. Biol.* **6**.

Inano, K., Suetake, I., Ueda, T., Miyake, Y., Nakamura, M., Okada, M., and Tajima, S. (2000) Maintenance-type DNA methyltransferase is highly expressed in post-mitotic neurons and localized in the cytoplasmic compartment. *J. Biochem. (Tokyo)* **128,** 315-321.

Ivanova, N. B., Dimos, J. T., Schaniel, C., Hackney, J. A., Moore, K. A., and Lemischka, I. R. (2002) A stem cell molecular signature. *Science* **298,** 601-604.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., and Jaenisch, R. (2001) Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.* **27,** 31-39.

Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004) Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. *Mol Cell Biol* **24,** 8862-8871.

Jaenisch, R. and Bird, A. (2003) Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat. Genet.* **33 Suppl,** 245-254.

Jair, K. W., Bachman, K. E., Suzuki, H., Ting, A. H., Rhee, I., Yen, R. W. C., Baylin, S. B., and Schuebel, K. E. (2006) De novo CpG island methylation in human cancer cells. *Cancer Research* **66,** 682-692.

Janicki, S. M., Tsukamoto, T., Salghetti, S. E., Tansey, W. P., Sachidanandam, R., Prasanth, K. V., Ried, T., Shav-Tal, Y., Bertrand, E., Singer, R. H., and Spector, D. L. (2004) From silencing to gene expression: Real-time analysis in single cells. *Cell* **116,** 683-698.

Jeltsch, A. (2006) On the Enzymatic Properties of Dnmt1 Specificity, Processivity, Mechanism of Linear Diffusion and Allosteric Regulation of the Enzyme. *Epigenetics* **1,** 63-66.

Jenuwein, T. and Allis, C. D. (2001) Translating the histone code. *Science* **293,** 1074-1080.

Jia, D., Jurkowska, R. Z., Zhang, X., Jeltsch, A., and Cheng, X. D. (2007) Structure of Dnmt3a bound to Dnmt3L suggests a model for de novo DNA methylation. *Nature* **449,** 248-U13.

Jiang, C. L., Jin, S. G., and Pfeifer, G. P. (2004) MBD3L1 is a transcriptional repressor that interacts with methyl-CpG-binding protein 2 (MBD2) and components of the NuRD complex. *J. Biol. Chem.* **279,** 52456-52464.

Jin, C. Y. and Felsenfeld, G. (2007) Nucleosome stability mediated by histone variants H3.3 and H2A.Z. *Genes Dev* **21,** 1519-1529.

Jiricny, J. (1998) Replication errors: cha(lle)nging the genome. *EMBO J.* **17,** 6427-6436.

Johnson, A. D., Crother, B., White, M. E., Patient, R., Bachvarova, R. F., Drum, M., and Masi, T. (2003) Regulative germ cell specification in axolotl embryos: a primitive trait conserved in the mammalian lineage. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **358,** 1371-1379.

Jones, P. A. and Laird, P. W. (1999) Cancer epigenetics comes of age. *Nat. Genet.* **21,** 163-167.

Jones, P. L., Veenstra, G. J., Wade, P. A., Vermaak, D., Kass, S. U., Landsberger, N., Strouboulis, J., and Wolffe, A. P. (1998) Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat. Genet.* **19,** 187-191.

Jorgensen, H. F., Ben Porath, I., and Bird, A. P. (2004) Mbd1 is recruited to both methylated and nonmethylated CpGs via distinct DNA binding domains. *Mol Cell Biol* **24,** 3387-3395.

Jorgensen, H. F., Adie, K., Chaubert, P., and Bird, A. P. (2006) Engineering a high-affinity methyl-CpG-binding protein. *Nucleic Acids Res.* **34,** e96.

Joshi, A. A. and Struhl, K. (2005) Eaf3 chromodomain interaction with methylated H3-K36 links histone deacetylation to Pol II elongation. *Mol. Cell* **20,** 971-978.

Kaminsky, Z. A., Tang, T., Wang, S. C., Ptak, C., Oh, G. H. T., Wong, A. H. C., Feldcamp, L. A., Virtanen, C., Halfvarson, J., Tysk, C., Mcrae, A. F., Visscher, P. M., Montgomery, G. W., Gottesman, I. I., Martin, N. G., and Petronis, A. (2009) DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genetics* **41,** 240-245.

Kanduri, C., Thakur, N., and Pandey, R. R. (2006) The length of the transcript encoded from the Kcnq1ot1 antisense promoter determines the degree of silencing. *EMBO Journal* **25,** 2096-2106.

Kaneda, M., Okano, M., Hata, K., Sado, T., Tsujimoto, N., Li, E., and Sasaki, H. (2004) Essential role for de novo DNA methyltransferase Dnmt3a in paternal and maternal imprinting. *Nature* **429,** 900-903.

Karpen, G. H. and Allshire, R. C. (1997) The case for epigenetic effects on centromere identity and function. *Trends Genet.* **13,** 489-496.

Kass, S. U., Landsberger, N., and Wolffe, A. P. (1997) DNA methylation directs a time dependent repression of transcription initiation. *Current Biology* **7,** 157-165.

Kato, Y. and Sasaki, H. (2005) Imprinting and looping: epigenetic marks control interactions between regulatory elements. *Bioessays* **27,** 1-4.

Kawai, J., Hirose, K., Fushiki, S., Hirotsune, S., Ozawa, N., Hara, A., Hayashizaki, Y., and Watanabe, S. (1994) Comparison of DNA methylation patterns among mouse cell lines by restriction landmark genomic scanning. *Mol. Cell Biol.* **14,** 7421-7427.

Keshet, I., Schlesinger, Y., Farkash, S., Rand, E., Hecht, M., Segal, E., Pikarski, E., Young, R. A., Niveleau, A., Cedar, H., and Simon, I. (2006) Evidence for an instructive mechanism of de novo methylation in cancer cells. *Nat. Genet.* **38,** 149-153.

Khorasanizadeh, S. (2004) The nucleosome: from genomic organization to genomic regulation. *Cell* **116,** 259-272.

Kim, G. D., Ni, J., Kelesoglu, N., Roberts, R. J., and Pradhan, S. (2002) Co-operation and communication between the human maintenance and de novo DNA (cytosine-5) methyltransferases. *EMBO J.* **21,** 4183-4195.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S. H. (2008) An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132,** 1049-1061.

Kim, M., Trinh, B. N., Long, T. I., Oghamian, S., and Laird, P. W. (2004) Dnmt1 deficiency leads to enhanced microsatellite instability in mouse embryonic stem cells. *Nucleic Acids Res.* **32,** 5742-5749.

Kim, S., Park, J., Spring, C. M., Sater, A. K., Ji, H., Otchere, A. A., Daniel, J. M., and Mccrea, P. D. (2004) Non-canonical Wnt signals are modulated by the Kaiso transcriptional repressor and p120-catenin. *Molecular Biology of the Cell* **15,** 360A.

Kimura, H. and Shiota, K. (2003) Methyl-CpG-binding protein, MeCP2, is a target molecule for maintenance DNA methyltransferase, Dnmt1. *J. Biol. Chem.* **278,** 4806-4812.

Klattenhoff, C. and Theurkauf, W. (2008) Biogenesis and germline functions of piRNAs. *Development* **135,** 3-9.

Klattig, J., Sierig, R., Kruspe, D., Besenbeck, B., and Englert, C. (2007) Wilms' tumor protein Wt1 is an activator of the anti-Mullerian hormone receptor gene Amhr2. *Mol. Cell Biol.* **27,** 4355-4364.

Klimasauskas, S., Kumar, S., Roberts, R. J., and Cheng, X. D. (1994) Hhal Methyltransferase Flips Its Target Base Out of the DNA Helix. *Cell* **76,** 357-369.

Klose, R. J., Sarraf, S. A., Schmiedeberg, L., McDermott, S. M., Stancheva, I., and Bird, A. P. (2005) DNA binding selectivity of MeCP2 due to a requirement for A/T sequences adjacent to methyl-CpG. *Mol. Cell* **19,** 667-678.

Klose, R. J. and Bird, A. P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31,** 89-97.

Kockar, F. T., Foka, P., Hughes, T. R., Kousteni, S., and Ramji, D. P. (2001) Analysis of the Xenopus laevis CCAAT-enhancer binding protein alpha gene promoter demonstrates species-specific differences in the mechanisms for both auto-activation and regulation by Sp1. *Nucleic Acids Res.* **29,** 362-372.

Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004) A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *Plos Biology* **2,** 991-1003.

Kondo, E., Gu, Z., Horii, A., and Fukushige, S. (2005) The thymine DNA glycosylase MBD4 represses transcription and is associated with methylated p16(INK4a) and hMLH1 genes. *Mol. Cell Biol.* **25,** 4388-4396.

Kouzarides, T. (2007) Chromatin modifications and their function. *Cell* **128,** 693-705.

Kovesdi, I., Reichel, R., and Nevins, J. R. (1987) Role of An Adenovirus E2 Promoter Binding-Factor in E1A-Mediated Coordinate Gene-Control. *Proc. Natl. Acad. Sci. U.S.A.* **84,** 2180-2184.

Kriaucionis, S., Paterson, A., CurtiS, J., Guy, J., MacLeod, N., and Bird, A. (2006) Gene expression analysis exposes mitochondrial abnormalities in a mouse model of Rett syndrome. *Mol Cell Biol* **26,** 5033-5042.

Kriaucionis, S. and Heintz, N. (2009) The Nuclear DNA Base 5-Hydroxymethylcytosine Is Present in Purkinje Neurons and the Brain. *Science* **324,** 929-930.

Kubicek, S., O'Sullivan, R. J., August, E. M., Hickey, E. R., Zhan, Q., Teodoro, M. L., Rea, S., Mechtler, K., Kowalski, J. A., Homon, C. A., Kelly, T. A., and Jenuwein, T. (2007) Reversal of H3K9me2 by a small-molecule inhibitor for the G9a histone methyltransferase. *Mol. Cell* **25,** 473-481.

Kubota, H. and Brinster, R. L. (2006) Technology Insight: in vitro culture of spermatogonial stem cells and their potential therapeutic uses. *Nature Clinical Practice Endocrinology & Metabolism* **2,** 99-108.

Kucharski, R., Maleszka, J., Foret, S., and Maleszka, R. (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science* **319,** 1827-1830.

Kuntz, S., Kieffer, E., Bianchetti, L., Lamoureux, N., Fuhrmann, G., and Viville, S. (2008) Tex19, a mammalian-specific protein with a restricted expression in pluripotent stem cells and germ line. *Stem Cells* **26,** 734-744.

Kuramochi-Miyagawa, S., Kimura, T., Ijiri, T. W., Isobe, T., Asada, N., Fujita, Y., Ikawa, M., Iwai, N., Okabe, M., Deng, W., Lin, H. F., Matsuda, Y., and Nakano, T. (2004) Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131,** 839-849.

Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004) Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell* **14,** 183-193.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410,** 116-120.

Lagger, G., O'Carroll, D., REMBOld, M., Khier, H., Tischler, J., Weitzer, G., Schuettengruber, B., Hauser, C., Brunmeir, R., Jenuwein, T., and Seiser, C. (2002) Essential function of histone deacetylase 1 in proliferation control and CDK inhibitor repression. *EMBO J.* **21,** 2672-2681.

Lande-Diner, L. and Cedar, H. (2005) Silence of the genes - mechanisms of long-term repression. *Nature Reviews Genetics* **6,** 650-654.

Lande-Diner, L., Zhang, J., Ben Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A. G., Rechavi, G., and Cedar, H. (2007) Role of DNA methylation in stable gene repression. *J. Biol. Chem.* **282,** 12194-12200.

Lane, N., Dean, W., Erhardt, S., Hajkova, P., Surani, A., Walter, J., and Reik, W. (2003) Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35,** 88-93.

Lawlor, E. R., Soucek, L., Brown-Swigart, L., Shchors, K., Bialucha, C. U., and Evan, G. I. (2006) Reversible kinetic analysis of Myc targets in vivo provides novel insights into Myc-mediated tumorigenesis. *Cancer Research* **66,** 4591-4601.

Lawson, K. A. and Hage, W. J. (1994) Clonal Analysis of the Origin of Primordial Germ-Cells in the Mouse. *Germline Development* **182,** 68-84.

Lawson, K. A., Dunn, N. R., Roelen, B. A. J., Zeinstra, L. M., Davis, A. M., Wright, C. V. E., Korving, J. P. W. F., and Hogan, B. L. M. (1999) Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes Dev* **13,** 424-436.

Lee, A. P., Koh, E. G., Tay, A., Brenner, S., and Venkatesh, B. (2006) Highly conserved syntenic blocks at the vertebrate Hox loci and conserved regulatory elements within and outside Hox gene clusters. *Proc. Natl. Acad. Sci. U. S. A* **103,** 6994-6999.

Lee, J. H., Hart, S. R. L., and Skalnik, D. G. (2004) Histone deacetylase activity is required for embryonic stem cell differentiation. *Genesis* **38,** 32-38.

Lee, P. J., Washer, L. L., Law, D. J., Boland, C. R., Horon, I. L., and Feinberg, A. P. (1996) Limited up-regulation of DNA methyltransferase in human colon cancer reflecting increased cell proliferation. *Proc. Natl. Acad. Sci. U. S. A* **93,** 10366-10370.

Lee, P. P., Fitzpatrick, D. R., Beard, C., Jessup, H. K., Lehar, S., Makar, K. W., Perez-Melgosa, M., Sweetser, M. T., Schlissel, M. S., Nguyen, S., Cherry, S. R., Tsai, J. H., Tucker, S. M., Weaver, W. M., Kelso, A., Jaenisch, R., and Wilson, C. B. (2001) A critical role for Dnmt1 and DNA methylation in T cell development, function, and survival. *Immunity* **15,** 763-774.

Lei, H., Oh, S. P., Okano, M., Juttermann, R., Goss, K. A., Jaenisch, R., and Li, E. (1996) De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122,** 3195-3205.

Leonhardt, H., Page, A. W., Weier, H. U., and Bestor, T. H. (1992) A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei. *Cell* **71,** 865-873.

Lessard, J., Aicha, S. B., Fournier, A., Calvo, E., Lavergne, E., Pelletier, M., and Labrie, C. (2007) Characterization of the RSL1-dependent conditional expression system in LNCaP prostate cancer cells and development of a single vector format. *Prostate* **67,** 808-819.

Lewis, A., Mitsuya, K., Umlauf, D., Smith, P., Dean, W., Walter, J., Higgins, M., Feil, R., and Reik, W. (2004) Imprinting on distal chromosome 7 in the placenta involves repressive histone methylation independent of DNA methylation. *Nature Genetics* **36,** 1291-1295.

Lewis, A., Green, K., Dawson, C., Redrup, L., Huynh, K. D., Lee, J. T., Hemberger, M., and Reik, W. (2006) Epigenetic dynamics of the Kcnq1 imprinted domain in the early embryo. *Development* **133,** 4203-4210.

Lewis, J. D., Meehan, R. R., Henzel, W. J., Maurerfogy, I., Jeppesen, P., Klein, F., and Bird, A. (1992) Purification, Sequence, and Cellular-Localization of A Novel Chromosomal Protein That Binds to Methylated DNA. *Cell* **69,** 905-914.

Li, E., Bestor, T. H., and Jaenisch, R. (1992) Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69,** 915-926.

Li, E., Beard, C., and Jaenisch, R. (1993) Role for DNA methylation in genomic imprinting. *Nature* **366,** 362-365.

Li, E. (2002) Chromatin modification and epigenetic reprogramming in mammalian development. *Nat. Rev. Genet.* **3,** 662-673.

Li, J. W., Lin, Q. S., Wang, W. D., Wade, P., and Wong, J. M. (2002) Specific targeting and constitutive association of histone deacetylase complexes during transcriptional repression. *Genes Dev* **16,** 687-692.

Li, J. Y., Lees-Murdock, D. J., Xu, G. L., and Walsh, C. P. (2004) Timing of establishment of paternal methylation imprints in the mouse. *Genomics* **84,** 952-960.

Li, J. Y., Pu, M. T., Hirasawa, R., Li, B. Z., Huang, Y. N., Zeng, R., Jing, N. H., Chen, T., Li, E., Sasaki, H., and Xu, G. L. (2007) Synergistic function of DNA methyltransferases Dnmt3a and Dnmt3b in the methylation of Oct4 and Nanog. *Mol. Cell Biol.* **27,** 8748-8759.

Li, Y., Kirschmann, D. A., and Wallrath, L. L. (2002) Does heterochromatin protein 1 always follow code? *Proc. Natl. Acad. Sci. U. S. A* **99 Suppl 4,** 16462-16469.

Liang, G., Chan, M. F., Tomigahara, Y., Tsai, Y. C., Gonzales, F. A., Li, E., Laird, P. W., and Jones, P. A. (2002) Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol. Cell Biol.* **22,** 480-491.

Lister, R., Pelizzola, M., Dowen, R. H., Hawkins, R. D., Hon, G., Tonti-Filippini, J., Nery, J. R., Lee, L., Ye, Z., Ngo, Q. M., Edsall, L., Antosiewicz-Bourget, J., Stewart, R., Ruotti, V., Millar, A. H., Thomson, J. A., Ren, B., and Ecker, J. R. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315-322.

Liu, P., Jenkins, N. A., and Copeland, N. G. (2003) A highly efficient recombineering-based method for generating conditional knockout mutations. *Genome Res.* **13,** 476-484.

Lock, L. F., Takagi, N., and Martin, G. R. (1987) Methylation of the Hprt Gene on the Inactive-X Occurs After Chromosome Inactivation. *Cell* **48,** 39-46.

Loh, Y. H., Wu, Q., Chew, J. L., Vega, V. B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., Wong, K. Y., Sung, K. W., Lee, C. W., Zhao, X. D., Chiu, K. P., Lipovich, L., Kuznetsov, V. A., Robson, P., Stanton, L. W., Wei, C. L., Ruan, Y., Lim, B., and Ng, H. H. (2006) The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat. Genet.* **38,** 431-440.

Loh, Y. H., Zhang, W. W., Chen, X., George, J., and Ng, H. H. (2007) Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev* **21,** 2545-2557.

Lopez-Serra, L., Ballestar, E., Ropero, S., Setien, F., Billard, L. M., Fraga, M. F., Lopez-Nieva, P., Alaminos, M., Guerrero, D., Dante, R., and Esteller, M. (2008) Unmasking of epigenetically

silenced candidate tumor suppressor genes by removal of methyl-CpG-binding domain proteins. *Oncogene* **27,** 3556-3566.

Lovell-Badge, R. (2001) The future for stem cell research. *Nature* **414,** 88-91.

Lucifero, D., Mann, M. R. W., Bartolomei, M. S., and Trasler, J. M. (2004) Gene-specific timing and epigenetic memory in oocyte imprinting. *Hum. Mol. Gen* **13,** 839-849.

Luger, K., Mader, A. W., Richmond, R. K., Sargent, D. F., and Richmond, T. J. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251-260.

Lunyak, V. V., Prefontaine, G. G., Nunez, E., Cramer, T., Ju, B. G., Ohgi, K. A., Hutt, K., Roy, R., Garcia-Diaz, A., Zhu, X. Y., Yung, Y., Montoliu, L., Glass, C. K., and Rosenfeld, M. G. (2007) Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* **317,** 248-251.

Maatouk, D. M., Kellam, L. D., Mann, M. R., Lei, H., Li, E., Bartolomei, M. S., and Resnick, J. L. (2006) DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133,** 3411-3418.

MacLean, J.A. & Wilkinson, M.F. (2005) Gene Regulatory Mechanisms in the Testis. In: *"Current Trends in Developmental Biology"* **71**:131-197.

Maekawa, M., Yamamoto, T., Tanoue, T., Yuasa, Y., Chisaka, O., and Nishida, E. (2005) Requirement of the MAP kinase signaling pathways for mouse preimplantation development. *Development* **132,** 1773-1783.

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminli, S., Arnold, K., Stadtfeld, M., Yachechko, R., Tchieu, J., Jaenisch, R., Plath, K., and Hochedlinger, K. (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1,** 55-70.

Mahy, N. L., Perry, P. E., Gilchrist, S., Baldock, R. A., and Bickmore, W. A. (2002) Spatial organization of active and inactive genes and noncoding DNA within chromosome territories. *J. Cell Biol.* **157,** 579-589.

Maison, C. and Almouzni, G. (2004) HP1 and the dynamics of heterochromatin maintenance. *Nat. Rev. Mol. Cell Biol.* **5,** 296-304.

Majumder, S., Kutay, H., Datta, J., Summers, D., Jacob, S. T., and Ghoshal, K. (2006) Epigenetic regulation of metallothionein-1 gene expression: Differential regulation of methylated and unmethylated promoters by DNA methyltransferases and methyl CpG binding proteins. *Journal of Cellular Biochemistry* **97,** 1300-1316.

Mak, W., Nesterova, T. B., de Napoles, M., Appanah, R., Yamanaka, S., Otte, A. P., and Brockdorff, N. (2004) Reactivation of the paternal X chromosome in early mouse embryos. *Science* **303,** 666-669.

Malik, H. S., Burke, W. D., and Eickbush, T. H. (1999) The age and evolution of non-LTR retrotransposable elements. *Mol. Biol. Evol.* **16,** 793-805.

Malik, H. S. and Henikoff, S. (2003) Phylogenomics of the nucleosome. *Nat. Struct. Biol.* **10,** 882-891.

Maaloe, O and Kjeldgaard, N.O. (1966). *Control of Macromolecular Synthesis: A Study of DNA, RNA and Protein Synthesis in Bacteria*.New York press.

Margot, J. B., Aguirre-Arteta, A. M., Di Giacco, B. V., Pradhan, S., Roberts, R. J., Cardoso, M. C., and Leonhardt, H. (2000) Structure and function of the mouse DNA methyltransferase gene: Dnmt1 shows a tripartite structure. *J. Mol. Biol.* **297,** 293-300.

Margot, J. B., Ehrenhofer-Murray, A. E., and Leonhardt, H. (2003) Interactions within the mammalian DNA methyltransferase family. *BMC. Mol. Biol.* **4,** 7.

Martens, J. H., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., and Jenuwein, T. (2005) The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* **24,** 800-812.

Martin, C., I, Hansen, J., Leaford, D., Pollard, S., and Hendrich, B. D. (2009) The methyl-CpG binding proteins Mecp2, Mbd2 and Kaiso are dispensable for mouse embryogenesis, but play a redundant function in neural differentiation. *PLoS One.* **4,** e4315.

Martin, G. R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. U. S. A* **78,** 7634-7638.

Matsui, Y., Zsebo, K., and Hogan, B. L. (1992) Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70,** 841-847.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000) Demethylation of the zygotic paternal genome. *Nature* **403,** 501-502.

McCabe, M. T., Davis, J. N., and Day, M. L. (2005) Regulation of DNA methyltransferase 1 by the pRb/E2F1 pathway. *Cancer Res.* **65,** 3624-3632.

McCarrey, J. R., Geyer, C. B., and Yoshioka, H. (2005) Epigenetic regulation of testis-specific gene expression. *Testicular Cell Dynamics and Endocrine Signaling* **1061,** 226-242.

McGhee, J. D. and Felsenfeld, G. (1980) The number of charge-charge interactions stabilizing the ends of nucleosome DNA. *Nucleic Acids Res.* **8,** 2751-2769.

Mclaren, A. (1983) Germ-Cell Sex Determination. *American Journal of Human Genetics* **35,** A12.

Meehan, R. R., Lewis, J. D., Mckay, S., Kleiner, E. L., and Bird, A. P. (1989) Identification of A Mammalian Protein That Binds Specifically to DNA Containing Methylated CpGs. *Cell* **58,** 499-507.

Meehan, R. R., Lewis, J. D., and Bird, A. P. (1992) Characterization of Mecp2, A Vertebrate DNA-Binding Protein with Affinity for Methylated DNA. *Nucleic Acids Research* **20,** 5085-5092.

Meehan, R. R. and Stancheva, I. (2001) DNA methylation and control of gene expression in vertebrate development. *Essays Biochem.* **37,** 59-70.

Meehan, R. R. (2003) DNA methylation in animal development. *Semin. Cell Dev. Biol.* . **14,** 53-65.

Meehan, R. R., Dunican, D. S., Ruzov, A., and Pennings, S. (2005) Epigenetic silencing in embryogenesis. *Exp. Cell Res.* **309,** 241-249.

Meissner, A., Mikkelsen, T. S., Gu, H. C., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X. L., Bernstein, B. E., Nusbaum, C., Jaffe, D. B., Gnirke, A., Jaenisch, R., and Lander, E. S. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454,** 766-U91.

Mello, J. A. and Almouzni, G. (2001) The ins and outs of nucleosome assembly. *Curr. Opin. Genet. Dev.* **11,** 136-141.

Meneghini, M. D., Wu, M., and Madhani, H. D. (2003) Conserved histone variant H2A.Z protects euchromatin from the ectopic spread of silent heterochromatin. *Cell* **112,** 725-736.

Mersfelder, E. L. and Parthun, M. R. (2006) The tale beyond the tail: histone core domain modifications and the regulation of chromatin structure. *Nucleic Acids Res.* **34,** 2653-2662.

Mertineit, C., Yoder, J. A., Taketo, T., Laird, D. W., Trasler, J. M., and Bestor, T. H. (1998) Sex-specific exons control DNA methyltransferase in mammalian germ cells. *Development* **125,** 889-897.

Migeon, B. R., Debeur, S. J., and Axelman, J. (1989) Frequent Derepression of G6Pd and Hprt on the Marsupial Inactive X-Chromosome Associated with Cell-Proliferation Invitro. *Experimental Cell Research* **182,** 597-609.

Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T. K., Koche, R. P., Lee, W., Mendenhall, E., O'donovan, A., Presser, A., Russ, C., Xie, X., Meissner, A., Wernig, M., Jaenisch, R., Nusbaum, C., Lander, E. S., and Bernstein, B. E. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature*.

Mikkelsen, T. S., Hanna, J., Zhang, X. L., Ku, M. C., Wernig, M., Schorderet, P., Bernstein, B. E., Jaenisch, R., Lander, E. S., and Meissner, A. (2008) Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49-U1.

Millar, C. B., Guy, J., Sansom, O. J., Selfridge, J., MacDougall, E., Hendrich, B., Keightley, P. D., Bishop, S. M., Clarke, A. R., and Bird, A. (2002) Enhanced CpG mutability and tumorigenesis in MBD4-deficient mice. *Science* **297,** 403-405.

Milutinovic, S., Brown, S. E., Zhuang, Q. L., and Szyf, M. (2004) DNA methyltransferase 1 knock down induces gene expression by a mechanism independent of DNA methylation and histone deacetylation. *J Biol Chem.* **279,** 27915-27927.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113,** 631-642.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003) The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113,** 631-642.

Mohn, F., Weber, M., Rebhan, M., Roloff, T. C., Richter, J., Stadler, M. B., Bibel, M., and Schubeler, D. (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. *Mol. Cell* **30,** 755-766.

Mohn, F. and Schubeler, D. (2009) Genetics and epigenetics: stability and plasticity during cellular differentiation. *Trends in Genetics* **25,** 129-136.

Monk, M. and Mclaren, A. (1981) X-Chromosome Activity in Fetal Germ-Cells of the Mouse. *Journal of Embryology and Experimental Morphology* **63,** 75-84.

Montgomery, N. D., Yee, D., Chen, A., Kalantry, S., Chamberlain, S. J., Otte, A. P., and Magnuson, T. (2005) The murine polycomb group protein Eed is required for global histone H3 lysine-27 methylation. *Current Biology* **15,** 942-947.

Mortusewicz, O., Schermelleh, L., Walter, J., Cardoso, M. C., and Leonhardt, H. (2005) Recruitment of DNA methyltransferase I to DNA repair sites. *Proc. Natl. Acad. Sci. U. S. A* **102,** 8905-8909.

Muromoto, R., Sugiyama, K., Takachi, A., Imoto, S., Sato, N., Yamamoto, T., Oritani, K., Shimoda, K., and Matsuda, T. (2004) Physical and functional interactions between Daxx and DNA methyltransferase 1-associated protein, DMAP1. *J. Immunol.* **172,** 2985-2993.

Murrell, A., Heeson, S., and Reik, W. (2004) Interaction between differentially methylated regions partitions the imprinted genes Igf2 and H19 into parent-specific chromatin loops. *Nature Genetics* **36,** 889-893.

Nagai, M., Nakamura, A., Makino, R., and Mitamura, K. (2003) Expression of DNA (5-cytosin)-methyltransferases (DNMTs) in hepatocellular carcinomas. *Hepatol. Res.* **26,** 186-191.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* **26,** 101-106

Nakamura, T., Arai, Y., Umehara, H., Masuhara, M., Kimura, T., Taniguchi, H., Sekimoto, T., Ikawa, M., Yoneda, Y., Okabe, M., Tanaka, S., Shiota, K., and Nakano, T. (2007) PGC7/Stella protects against DNA demethylation in early embryogenesis. *Nature Cell Biology* **9,** 64-U81.

Nan, X., Ng, H. H., Johnson, C. A., Laherty, C. D., Turner, B. M., Eisenman, R. N., and Bird, A. (1998) Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex. *Nature* **393,** 386-389.

Nan, X. S., Meehan, R. R., and Bird, A. (1993) Dissection of the Methyl-CpG Binding Domain from the Chromosomal Protein Mecp2. *Nucleic Acids Research* **21,** 4886-4892.

Nan, X. S., Hou, J. H., Maclean, A., Nasir, J., Lafuente, M. J., Shu, X. H., Kriaucionis, S., and Bird, A. (2007) Interaction between chromatin proteins MECP2 and ATRX is disrupted by mutations that cause inherited mental retardation. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 2709-2714.

Nathan, D., Ingvarsdottir, K., Sterner, D. E., Bylebyl, G. R., Dokmanovic, M., Dorsey, J. A., Whelan, K. A., Krsmanovic, M., Lane, W. S., Meluh, P. B., Johnson, E. S., and Berger, S. L. (2006) Histone sumoylation is a negative regulator in Saccharomyces cerevisiae and shows dynamic interplay with positive-acting histone modifications. *Genes Dev.* **20,** 966-976.

Nelson, C. J., Santos-Rosa, H., and Kouzarides, T. (2006) Proline isomerization of histone H3 regulates lysine methylation and gene expression. *Cell* **126,** 905-916.

Ng, H. H., Zhang, Y., Hendrich, B., Johnson, C. A., Turner, B. M., Erdjument-Bromage, H., Tempst, P., Reinberg, D., and Bird, A. (1999) MBD2 is a transcriptional repressor belonging to the MeCP1 histone deacetylase complex. *Nature Genetics* **23,** 58-61.

Ng, H. H., Jeppesen, P., and Bird, A. (2000) Active repression of methylated genes by the chromosomal protein MBD1. *Mol Cell Biol* **20,** 1394-1406.

Ng, R. K. and Gurdon, J. B. (2008) Epigenetic memory of an active gene state depends on histone H3.3 incorporation into chromatin in the absence of transcription. *Nature Cell Biology* **10,** 102-U83.

Ng, R. K., Dean, W., Dawson, C., Lucifero, D., Madeja, Z., Reik, W., and Hemberger, M. (2008) Epigenetic restriction of embryonic cell lineage fate by methylation of Elf5. *Nature Cell Biology* **10,** 1280-1U68.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998) Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* **95,** 379-391.

Nieto, M., Samper, E., Fraga, M. F., Gonzalez, d. B., Esteller, M., and Serrano, M. (2004) The absence of p53 is critical for the induction of apoptosis by 5-aza-2'-deoxycytidine. *Oncogene* **23,** 735-743.

Nishikawa, S., Nishikawa, S., Hirashima, M., Matsuyoshi, N., and Kodama, H. (1998) Progressive lineage analysis by cell sorting and culture identifies FLK1(+)VE-cadherin(+) cells at a diverging point of endothelial and hemopoietic lineages. *Development* **125,** 1747-1757.

Nishikawa, S., Jakt, L. M., and Era, T. (2007) Embryonic stem-cell culture as a tool for developmental cell biology. *Nat. Rev. Mol. Cell Biol.* **8,** 502-507.

Niswander, L., Yee, D., Rinchik, E. M., Russell, L. B., and Magnuson, T. (1988) The Albino Deletion Complex and Early Postimplantation Survival in the Mouse. *Development* **102,** 45-53.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998) Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes Dev.* **12,** 2048-2060.

Noma, K. I., Cam, H. P., Maraia, R. J., and Grewal, S. I. S. (2006) A role for TFIIIC transcription factor complex in genome organization. *Cell* **125,** 859-872.

Norris, D. P., Brockdorff, N., and Rastan, S. (1991) Methylation Status of CpG-Rich Islands on Active and Inactive Mouse X-Chromosomes. *Mammalian Genome* **1,** 78-83.

O'Carroll, D., Erhardt, S., Pagani, M., Barton, S. C., Surani, M. A., and Jenuwein, T. (2001) The polycomb-group gene Ezh2 is required for early mouse development. *Mol. Cell Biol.* **21,** 4330-4336.

Oda, M., Yamagiwa, A., Yamamoto, S., Nakayama, T., Tsumura, A., Sasaki, H., Nakao, K., Li, E., and Okano, M. (2006) DNA methylation regulates long-range gene silencing of an X-linked homeobox gene cluster in a lineage-specific manner. *Genes Dev.* **20,** 3382-3394.

Ohinata, Y., Payer, B., O'Carroll, D., Ancelin, K., Ono, Y., Sano, M., Barton, S. C., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., Saitou, M., and Surani, M. A. (2005) Blimp1 is a critical determinant of the germ cell lineage in mice. *Nature* **436,** 207-213.

Ohki, I., Shimotake, N., Fujita, N., Jee, J. G., Ikegami, T., Nakao, M., and Shirakawa, M. (2001) Solution structure of the methyl-CpG binding domain of human MBD1 in complex with methylated DNA. *Cell* **105,** 487-497.

Ohsawa, K., Imai, Y., Ito, D., and Kohsaka, S. (1996) Molecular cloning and characterization of annexin V-binding proteins with highly hydrophilic peptide structure. *Journal of Neurochemistry* **67,** 89-97.

Oka, M., Meacham, A. M., Hamazaki, T., Rodic, N., Chang, L. J., and Terada, N. (2005) De novo DNA methyltransferases Dnmt3a and Dnmt3b primarily mediate the cytotoxic effect of 5-aza-2'-deoxycytidine. *Oncogene* **24,** 3091-3099.

Okabe, S., Forsberg-Nilsson, K., Spiro, A. C., Segal, M., and McKay, R. D. (1996) Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. *Mech. Dev.* **59,** 89-102.

Okamoto, I., Arnaud, D., Le Baccon, P., Otte, A. P., Disteche, C. M., Avner, P., and Heard, E. (2005) Evidence for de novo imprinted X-chromosome inactivation independent of meiotic inactivation in mice. *Nature* **438,** 369-373.

Okano, M., Xie, S., and Li, E. (1998) Dnmt2 is not required for de novo and maintenance methylation of viral DNA in embryonic stem cells. *Nucleic Acids Res.* **26,** 2536-2540.

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999) DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99,** 247-257.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007) Generation of germline-competent induced pluripotent stem cells. *Nature* **448,** 313-317.

Olins, A. L. and Olins, D. E. (1974) Spheroid chromatin units (v bodies). *Science* **183,** 330-332.

Ollinger, R., Childs, A. J., Burgess, H. M., Speed, R. M., Lundegaard, P. R., Reynolds, N., Gray, N. K., Cooke, H. J., and Adams, I. R. (2008) Deletion of the Pluripotency-Associated Tex19.1 Gene Causes Activation of Endogenous Retroviruses and Defective Spermatogenesis in Mice. *Plos Genetics* **4**.

Ooi, S. K. T., Qiu, C., Bernstein, E., Li, K. Q., Jia, D., Yang, Z., Erdjument-Bromage, H., Tempst, P., Lin, S. P., Allis, C. D., Cheng, X. D., and Bestor, T. H. (2007) DNMT3L connects unmethylated lysine 4 of histone H3 to de novo methylation of DNA. *Nature* **448,** 714-U13.

Orford, K., Kharchenko, P., Lai, W., Dao, M. C., Worhunsky, D. J., Ferro, A., Janzen, V., Park, P. J., and Scadden, D. T. (2008) Differential H3K4 methylation identifies developmentally poised hematopoietic genes. *Dev. Cell* **14,** 798-809.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000) Active demethylation of the paternal genome in the mouse zygote. *Current Biology* **10,** 475-478.

Palli, S. R., Kapitskaya, M. Z., Kumar, M. B., and Cress, D. E. (2003) Improved ecdysone receptor-based inducible gene regulation system. *Eur. J. Biochem.* **270,** 1308-1315.

Palmer, D. K., Oday, K., Trong, H. L., Charbonneau, H., and Margolis, R. L. (1991) Purification of the Centromere-Specific Protein Cenp-A and Demonstration That It Is A Distinctive Histone. *Proc. Natl. Acad. Sci. U.S.A.* **88,** 3734-3738.

Pan, G. J., Tian, S. L., Nie, J., Yang, C. H., Ruotti, V., Wei, H. R., Jonsdottir, G. A., Stewart, R., and Thomson, J. A. (2007) Whole-genome analysis of histone H3 lysine 4 and lysine 27 methylation in human embryonic stem cells. *Cell Stem Cell* **1,** 299-312.

Panning, B. and Jaenisch, R. (1996) DNA hypomethylation can activate Xist expression and silence X-linked genes. *Genes Dev.* **10,** 1991-2002.

Patra, S. K., Patra, A., Zhao, H., and Dahiya, R. (2002) DNA methyltransferase and demethylase in human prostate cancer. *Mol. Carcinog.* **33,** 163-171.

Perez-Iratxeta, C., Palidwor, G., Porter, C. J., Sanche, N. A., Huska, M. R., Suomela, B. P., Muro, E. M., Krzyzanowski, P. M., Hughes, E., Campbell, P. A., Rudnicki, M. A., and Andrade, M. A. (2005) Study of stem cell function using microarray experiments. *FEBS Lett.* **579,** 1795-1801.

Peters, A. H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., Weipoltshammer, K., Pagani, M., Lachner, M., Kohlmaier, A., Opravil, S., Doyle, M., Sibilia, M., and Jenuwein, T. (2001) Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107,** 323-337.

Phalke, S., Nickel, O., Walluscheck, D., Hortig, F., Onorati, M. C., and Reuter, G. (2009) Retrotransposon silencing and telomere integrity in somatic cells of Drosophila depends on the cytosine-5 methyltransferase DNMT2. *Nature Genetics* **41,** 696-702.

Plass, C. and Soloway, P. D. (2002) DNA methylation, imprinting and cancer. *Eur. J. Hum. Genet.* **10,** 6-16.

Pohlers, M., Truss, M., Frede, U., Scholz, A., Strehle, M., Kuban, R. J., Hoffmann, B., Morkel, M., Birchmeier, C., and Hagemeier, C. (2005) A role for E2F6 in the restriction of male-germ-cell-specific gene expression. *Current Biology* **15,** 1051-1057.

Posfai, J., Bhagwat, A. S., Posfai, G., and Roberts, R. J. (1989) Predictive motifs derived from cytosine methyltransferases. *Nucleic Acids Res.* **17,** 2421-2435.

Potter, S. S., Hartman, H. A., Kwan, K. M., Behringer, R. R., and Patterson, L. T. (2007) Laser capture-microarray analysis of Lim1 mutant kidney development. *Genesis.* **45,** 432-439.

Pradhan, S. and Roberts, R. J. (2000) Hybrid mouse-prokaryotic DNA (cytosine-5) methyltransferases retain the specificity of the parental C-terminal domain. *EMBO Journal* **19,** 2103-2114.

Proffitt, J. H., Davie, J. R., Swinton, D., and Hattman, S. (1984) 5-Methylcytosine Is Not Detectable in Saccharomyces-Cerevisiae DNA. *Mol Cell Biol* **4,** 985-988.

Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A., and Prokhortchouk, E. (2001) The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev* **15,** 1613-1618.

Prokhortchouk, A., Sansom, O., Selfridge, J., Caballero, I. M., Salozhin, S., Aithozhina, D., Cerchietti, L., Meng, F. G., Augenlicht, L. H., Mariadason, J. M., Hendrich, B., Melnick, A., Prokhortchouk, E., Clarke, A., and Bird, A. (2006) Kaiso-deficient mice show resistance to intestinal cancer. *Mol. Cell Biol.* **26,** 199-208.

Pusarla, R. H. and Bhargava, P. (2005) Histones in functional diversification. Core histone variants. *FEBS J.* **272,** 5149-5168.

Qin, T., Jelinek, J., Si, J., Shu, J., and Issa, J. P. (2009) Mechanisms of resistance to 5-aza-2'-deoxycytidine in human cancer cell lines. *Blood* **113,** 659-667.

Qiu, C., Sawada, K., Zhang, X., and Cheng, X. (2002) The PWWP domain of mammalian DNA methyltransferase Dnmt3b defines a new family of DNA-binding folds. *Nat. Struct. Biol.* **9,** 217-224.

Rabinowicz, P. D., Palmer, L. E., May, B. P., Hemann, M. T., Lowe, S. W., McCombie, W. R., and Martienssen, R. A. (2003) Genes and transposons are differentially methylated in plants, but not in mammals. *Genome Research* **13,** 2658-2664.

Raisner, R. M. and Madhani, H. D. (2006) Patterning chromatin: form and function for H2A.Z variant nucleosomes. *Current Opinion in Genetics & Development* **16,** 119-124.

Rakyan, V. K., Hildmann, T., Novik, K. L., Lewin, J., Tost, J., Cox, A. V., Andrews, T. D., Howe, K. L., Otto, T., Olek, A., Fischer, J., Gut, I. G., Berlin, K., and Beck, S. (2004) DNA methylation profiling of the human major histocompatibility complex: A pilot study for the Human Epigenome Project. *Plos Biology* **2,** 2170-2182.

Ramirez-Carrozzi, V. R., Braas, D., Bhatt, D. M., Cheng, C. S., Hong, C., Doty, K. R., Black, J. C., Hoffmann, A., Carey, M., and Smale, S. T. (2009) A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138,** 114-128.

Ramsahoye, B. H., Biniszkiewicz, D., Lyko, F., Clark, V., Bird, A. P., and Jaenisch, R. (2000) Non-CpG methylation is prevalent in embryonic stem cells and may be mediated by DNA methyltransferase 3a. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 5237-5242.

Razin, A. and Cedar, H. (1977) Distribution of 5-methylcytosine in chromatin. *Proc. Natl. Acad. Sci. U. S. A* **74,** 2725-2728.

Razin, A. and Riggs, A. D. (1980) DNA Methylation and Gene-Function. *Science* **210,** 604-610.

Rea, S., Eisenhaber, F., O'Carroll, D., Strahl, B. D., Sun, Z. W., Schmid, M., Opravil, S., Mechtler, K., Ponting, C. P., Allis, C. D., and Jenuwein, T. (2000) Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406,** 593-599.

Reik, W. (2007) Stability and flexibility of epigenetic gene regulation in mammalian development. *Nature* **447,** 425-432.

Renda, M., Baglivo, I., Burgess-Beusse, B., Esposito, S., Fattorusso, R., Felsenfeld, G., and Pedone, P. V. (2007) Critical DNA binding interactions of the insulator protein CTCF - A small number of zinc fingers mediate strong binding, and a single finger-DNA interaction controls binding at imprinted loci. *J Biol Chem.* **282,** 33336-33345.

Riggs, A. D. & Porter, T. N. (1996) Overview of Epigenetic Mechanisms. *In Epigenetic mechanisms of gene regulation* (ed. W. E. A Russo, R. A. Martienseen & A. D. Riggs). Pp 28-41.
Rhee, I., Bachman, K. E., Park, B. H., Jair, K. W., Yen, R. W., Schuebel, K. E., Cui, H., Feinberg, A. P., Lengauer, C., Kinzler, K. W., Baylin, S. B., and Vogelstein, B. (2002) DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416,** 552-556.

Rietveld, L. E., Caldenhoven, E., and Stunnenberg, H. G. (2002) In vivo repression of an erythroid-specific gene by distinct corepressor complexes. *EMBO J.* **21,** 1389-1397.

Riggs, A. D. (1975) X-Inactivation, Differentiation, and DNA Methylation. *Cytogenetics and Cell Genetics* **14,** 9-25.

Ro, S., Park, C., Sanders, K. M., McCarrey, J. R., and Yan, W. (2007) Cloning and expression profiling of testis-expressed microRNAs. *Dev. Biol.* **311,** 592-602.

Robertson, K. D., Ait-Si-Ali, S., Yokochi, T., Wade, P. A., Jones, P. L., and Wolffe, A. P. (2000) DNMT1 forms a complex with Rb, E2F1 and HDAC1 and represses transcription from E2F-responsive promoters. *Nat. Genet.* **25,** 338-342.

Robinson, P. J. and Rhodes, D. (2006) Structure of the '30 nm' chromatin fibre: a key role for the linker histone. *Curr. Opin. Struct. Biol.* **16,** 336-343.

Rodda, D. J., Chew, J. L., Lim, L. H., Loh, Y. H., Wang, B., Ng, H. H., and Robson, P. (2005) Transcriptional regulation of Nanog by Oct4 and Sox2. *J Biol Chem.* **280,** 24731-24737.

Rodic, N., Oka, M., Hamazaki, T., Murawski, M. R., Jorgensen, M., Maatouk, D. M., Resnick, J. L., Li, E., and Terada, N. (2005) DNA methylation is required for silencing of Ant4, an adenine nucleotide translocase selectively expressed in mouse embryonic stem cells and germ cells. *Stem Cells* **23,** 1314-1323.

Roh, T. Y., Cuddapah, S., and Zhao, K. (2005) Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. *Genes Dev.* **19,** 542-552.

Roh, T. Y., Cuddapah, S., Cui, K., and Zhao, K. (2006) The genomic landscape of histone modifications in human T cells. *Proc. Natl. Acad. Sci. U. S. A* **103,** 15782-15787.

Rohwedel, J., Guan, K., and Wobus, A. M. (1999) Induction of cellular differentiation by retinoic acid in vitro. *Cells Tissues. Organs* **165,** 190-202.

Rolletschek, A., Chang, H., Guan, K., Czyz, J., Meyer, M., and Wobus, A. M. (2001) Differentiation of embryonic stem cell-derived dopaminergic neurons is enhanced by survival-promoting factors. *Mech. Dev.* **105,** 93-104.

Rossant, J. and McBurney, M. W. (1982) The developmental potential of a euploid male teratocarcinoma cell line after blastocyst injection. *J. Embryol. Exp. Morphol.* **70,** 99-112.

Rountree, M. R., Bachman, K. E., and Baylin, S. B. (2000) DNMT1 binds HDAC2 and a new co-repressor, DMAP1, to form a complex at replication foci. *Nat. Genet.* **25,** 269-277.

Ruggiu, M., Speed, R., Taggart, M., Mckay, S. J., Kilanowski, F., Saunders, P., Dorin, J., and Cooke, H. J. (1997) The mouse Dazla gene encodes a cytoplasmic protein essential for gametogenesis. *Nature* **389,** 73-77.

Ruzov, A., Dunican, D. S., Prokhortchouk, A., Pennings, S., Stancheva, I., Prokhortchouk, E., and Meehan, R. R. (2004) Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. *Development* **131,** 6185-6194.

Ruzov, A., Savitskaya, E., Hackett, J. A., Reddington, J. P., Prokhortchouk, A., Madej, M. J., Chekanov, N., Li, M. H., Dunican, D. S., Prokhortchouk, E., Pennings, S., and Meehan, R. R. (2009a) The non-methylated DNA-binding function of Kaiso is not required in early Xenopus laevis development. *Development* **136,** 729-738.

Ruzov, A., Hackett, J. A., Prokhortchouk, A., Reddington, J. P., Madej, M. J., Dunican, D. S., Prokhortchouk, E., Pennings, S., and Meehan, R. R. (2009b) The interaction of xKaiso with xTcf3: a revised model for integration of epigenetic and Wnt signalling pathways. *Development* **136,** 723-727.

Ruzov, A., Shorning, B., Mortusewicz, O., Dunican, D. S., Leonhardt, H., and Meehan, R. R. (2009c) MBD4 and MLH1 are required for apoptotic induction in xDNMT1-depleted embryos. *Development* **136,** 2277-2286.

Sado, T., Fenner, M. H., Tan, S. S., Tam, P., Shioda, T., and Li, E. (2000) X inactivation in the mouse embryo deficient for Dnmt1: Distinct effect of hypomethylation on imprinted and random X inactivation. *Dev. Biol.* **225,** 294-303.

Saeki, H., Ohsumi, K., Aihara, H., Ito, T., Hirose, S., Ura, K., and Kaneda, Y. (2005) Linker histone variants control chromatin dynamics during early embryogenesis. *Proc. Natl. Acad. Sci. U. S. A* **102,** 5697-5702.

Saito, M. and Ishikawa, F. (2002) The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J Biol Chem.* **277,** 35434-35439.

Saitou, M., Barton, S. C., and Surani, M. A. (2002) A molecular programme for the specification of germ cell fate in mice. *Nature* **418,** 293-300.

Samollow, P. B., Robinson, E. S., Ford, A. L., and Vandeberg, J. L. (1995) Developmental Progression of Gpd Expression from the Inactive X-Chromosome of the Virginia Opossum. *Developmental Genetics* **16,** 367-378.

Sanchez, C., Sanchez, I., Demmers, J. A. A., Rodriguez, P., Strouboulis, J., and Vidal, M. (2007) Proteomics analysis of Ring1B/Rnf2 interactors identifies a novel complex with the Fbxl10/Jhdm1B histone demethylase and the Bcl6 interacting corepressor. *Molecular & Cellular Proteomics* **6,** 820-834.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002) Dynamic reprogramming of DNA methylation in the early mouse embryo. *Dev. Biol.* . **241,** 172-182.

Sarraf, S. A. and Stancheva, I. (2004) Methyl-CpG binding protein MBD1 couples histone H3 methylation at lysine 9 by SETDB1 to DNA replication and chromatin assembly. *Mol. Cell* **15,** 595-605.

Sasai, N. and Defossez, P. A. (2009) Many paths to one goal? The proteins that recognize methylated DNA in eukaryotes. *International Journal of Dev. Biol.* **53,** 323-334.

Sasaki, H. and Matsui, Y. (2008) Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nature Reviews Genetics* **9,** 129-140.

Scanlan, M. J., Simpson, A. J., and Old, L. J. (2004) The cancer/testis genes: review, standardization, and commentary. *Cancer Immun.* **4,** 1.

Scheele, C., Petrovic, N., Faghihi, M. A., Lassmann, T., Fredriksson, K., Rooyackers, O., Wahlestedt, C., Good, L., and Timmons, J. A. (2007) The human PINK1 locus is regulated in vivo by a non-coding natural antisense RNA during modulation of mitochondrial function. *Bmc Genomics* **8**.

Schermelleh, L., Spada, F., Easwaran, H. P., Zolghadr, K., Margot, J. B., Cardoso, M. C., and Leonhardt, H. (2005) Trapped in action: direct visualization of DNA methyltransferase activity in living cells. *Nat. Methods* **2,** 751-756.

Schermelleh, L., Haernmer, A., Spada, F., Rosing, N., Meilinger, D., Rothbauer, U., Cardoso, M. C., and Leonhardt, H. (2007) Dynamics of Dnmt1 interaction with the replication machinery and

its role in postreplicative maintenance of DNA methylation. *Nucleic Acids Research* **35,** 4301-4312.

Schlesinger, Y., Straussman, R., Keshet, I., Farkash, S., Hecht, M., Zimmerman, J., Eden, E., Yakhini, Z., Ben Shushan, E., Reubinoff, B. E., Bergman, Y., Simon, I., and Cedar, H. (2007) Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nature Genetics* **39,** 232-236.

Schorderet, D. F. and Gartler, S. M. (1992) Analysis of CpG suppression in methylated and nonmethylated species. *Proc. Natl. Acad. Sci. U. S. A* **89,** 957-961.

Schotta, G., Lachner, M., Sarma, K., Ebert, A., Sengupta, R., Reuter, G., Reinberg, D., and Jenuwein, T. (2004) A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin. *Genes Dev.* **18,** 1251-1262.

Schreiber, S. L. and Bernstein, B. E. (2002) Signaling network model of chromatin. *Cell* **111,** 771-778.

Schuebel, K. E., Chen, W., Cope, L., Glockner, S. C., Suzuki, H., Yi, J. M., Chan, T. A., Van Neste, L., Van Criekinge, W., van den, B. S., van Engeland, M., Ting, A. H., Jair, K., Yu, W., Toyota, M., Imai, K., Ahuja, N., Herman, J. G., and Baylin, S. B. (2007) Comparing the DNA hypermethylome with gene mutations in human colorectal cancer. *PLoS Genet.* **3,** 1709-1723.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007) Genome regulation by polycomb and trithorax proteins. *Cell* **128,** 735-745.

Schwartz, Y. B. and Pirrotta, V. (2008) Polycomb complexes and epigenetic states. *Current Opinion in Cell Biology* **20,** 266-273.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thastrom, A., Field, Y., Moore, I. K., Wang, J. P., and Widom, J. (2006) A genomic code for nucleosome positioning. *Nature* **442,** 772-778.

Seki, Y., Hayashi, K., Itoh, K., Mizugaki, M., Saitou, M., and Matsui, Y. (2005) Extensive and orderly reprogramming of genome-wide chromatin modifications associated with specification and early development of germ cells in mice. *Dev. Biol.* **278,** 440-458.

Seki, Y., Yamaji, M., Yabuta, Y., Sano, M., Shigeta, M., Matsui, Y., Saga, Y., Tachibana, M., Shinkai, Y., and Saitou, M. (2007) Cellular dynamics associated with the genome-wide epigenetic reprogramming in migrating primordial germ cells in mice. *Development* **134,** 2627-2638.

Selker, E. U., Tountas, N. A., Cross, S. H., Margolin, B. S., Murphy, J. G., Bird, A. P., and Freitag, M. (2003) The methylated component of the Neurospora crassa genome. *Nature* **422,** 893-897.

Seydoux, G. and Braun, R. E. (2006) Pathway to totipotency: Lessons from germ cells. *Cell* **127,** 891-904.

Shamblott, M. J., Axelman, J., Wang, S. P., Bugg, E. M., Littlefield, J. W., Donovan, P. J., Blumenthal, P. D., Huggins, G. R., and Gearhart, J. D. (1998) Derivation of pluripotent stem cells horn cultured human primordial germ cells. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 13726-13731.

Sharif, J., Muto, M., Takebayashi, S. I., Suetake, I., Iwamatsu, A., Endo, T. A., Shinga, J., Mizutani-Koseki, Y., Toyoda, T., Okamura, K., Tajima, S., Mitsuya, K., Okano, M., and Koseki,

H. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* **450,** 908-U25.

Sharov, A. A., Masui, S., Sharova, L. V., Piao, Y., Aiba, K., Matoba, R., Xin, L., Niwa, H., and Ko, M. S. H. (2008) Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. *Bmc Genomics* **9**.

Shen, L., Kondo, Y., Guo, Y., Zhang, J., Zhang, L., Ahmed, S., Shu, J., Chen, X., Waterland, R. A., and Issa, J. P. (2007) Genome-wide profiling of DNA methylation reveals a class of normally methylated CpG island promoters. *PLoS Genet.* **3,** 2023-2036.

Shen, Y., Matsuno, Y., Fouse, S. D., Rao, N., Root, S., Xu, R. H., Pellegrini, M., Riggs, A. D., and Fan, G. P. (2008) X-inactivation in female human embryonic stem cells is in a nonrandom pattern and prone to epigenetic alterations. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 4709-4714.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J. R., Cole, P. A., Casero, R. A., and Shi, Y. (2004) Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119,** 941-953.

Shi, Y., Desponts, C., Do, J. T., Hahm, H. S., Scholer, H. R., and Ding, S. (2008) Induction of Pluripotent Stem Cells from Mouse Embryonic Fibroblasts by Oct4 and Klf4 with Small-Molecule Compounds. *Cell Stem Cell* **3,** 568-574.

Shin, S. I., Freedman, V. H., Risser, R., and Pollack, R. (1975) Tumorigenicity of virus-transformed cells in nude mice is correlated specifically with anchorage independent growth in vitro. *Proc. Natl. Acad. Sci. U. S A* **72,** 4435-4439.

Shiota, K., Kogo, Y., Ohgane, J., Imamura, T., Urano, A., Nishino, K., Tanaka, S., and Hattori, N. (2002) Epigenetic marks by DNA methylation specific to stem, germ and somatic cells in mice. *Genes to Cells* **7,** 961-969.

Siedlecki, P. and Zielenkiewicz, P. (2006) Mammalian DNA methyltransferases. *Acta Biochim. Pol.* **53,** 245-256.

Silva, C., Wood, J. R., Salvador, L., Zhang, Z. B., Kostetskii, I., Williams, C. J., and Strauss, J. F. (2009) Expression Profile of Male Germ Cell-Associated Genes in Mouse Embryonic Stem Cell Cultures Treated With All-Trans Retinoic Acid and Testosterone. *Molecular Reproduction and Development* **76,** 11-21.

Sing, A., Pannell, D., Karaiskakis, A., Sturgeon, K., Djabali, M., Ellis, J., Lipshitz, H. D., and Cordes, S. P. (2009) A Vertebrate Polycomb Response Element Governs Segmentation of the Posterior Hindbrain. *Cell* **138,** 885-897.

Small, C. L., Shima, J. E., Uzumcu, M., Skinner, M. K., and Griswold, M. D. (2005) Profiling gene expression during the differentiation and development of the murine embryonic gonad. *Biology of Reproduction* **72,** 492-501.

Smallwood, A., Esteve, P. O., Pradhan, S., and Carey, M. (2007) Functional cooperation between HP1 and DNMT1 mediates gene silencing. *Genes Dev.* **21,** 1169-1178.

Smit, A. F. and Riggs, A. D. (1996) Tiggers and DNA transposon fossils in the human genome. *Proc. Natl. Acad. Sci. U. S. A* **93,** 1443-1448.

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., and Rogers, D. (1988) Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. *Nature* **336,** 688-690.

Soyal, S. M., Amleh, A., and Dean, J. (2000) FIG alpha, a germ cell-specific transcription factor required for ovarian follicle formation. *Development* **127,** 4645-4654.

Spada, F., Haemmer, A., Kuch, D., Rothbauer, U., Schermelleh, L., Kremmer, E., Carell, T., Langst, G., and Leonhardt, H. (2007) DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J. Cell Biol.* **176,** 565-571.

Spivakov, M. and Fisher, A. G. (2007) Epigenetic signatures of stem-cell identity. *Nature Reviews Genetics* **8,** 263-271.

Spring, C. M., Kelly, K. F., O'Kelly, I., Graham, M., Crawford, H. C., and Daniel, J. M. (2005) The catenin p120(ctn) inhibits Kaiso-mediated transcriptional repression of the beta-catenin/TCF target gene matrilysin. *Experimental Cell Research* **305,** 253-265.

Stancheva, I. and Meehan, R. R. (2000) Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos. *Genes Dev.* **14,** 313-327.

Stancheva, I., Hensey, C., and Meehan, R. R. (2001) Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in Xenopus embryos. *EMBO J.* **20,** 1963-1973.

Stancheva, I., El Maarri, O., Walter, J., Niveleau, A., and Meehan, R. R. (2002) DNA methylation at promoter regions regulates the timing of gene activation in Xenopus laevis embryos. *Dev. Biol.* . **243,** 155-165.

Strahl, B. D., Ohba, R., Cook, R. G., and Allis, C. D. (1999) Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proc. Natl. Acad. Sci. U. S. A* **96,** 14967-14972.

Strahl, B. D. and Allis, C. D. (2000) The language of covalent histone modifications. *Nature* **403,** 41-45.

Straussman, R., Nejman, D., Roberts, D., Steinfeld, I., Blum, B., Benvenisty, N., Simon, I., Yakhini, Z., and Cedar, H. (2009) Developmental programming of CpG island methylation profiles in the human genome. *Nature Structural & Molecular Biology* **16,** 564-571.

Su, I., Dobenecker, M. W., Dickinson, E., Oser, M., Basavaraj, A., Marqueron, R., Viale, A., Reinberg, D., Wulfing, C., and Tarakhovsky, A. (2005) Polycomb group protein Ezh2 controls actin polymerization and cell signaling. *Cell* **121,** 425-436.

Suzuki, M., Sato, S., Arai, Y., Shinohara, T., Tanaka, S., Greally, J. M., Hattori, N., and Shiota, K. (2007) A new class of tissue-specifically methylated regions involving entire CpG islands in the mouse. *Genes Cells* **12,** 1305-1314.

Suzuki, M. M. and Bird, A. (2008) DNA methylation landscapes: provocative insights from epigenomics. *Nature Reviews Genetics* **9,** 465-476.

Svingen, T., Beverdam, A., Bernard, P., McClive, P., Harley, V. R., Sinclair, A. H., and Koopman, P. (2007) Sex-specific expression of a novel gene Tmem184a during mouse testis differentiation. *Reproduction* **133,** 983-989.

Szabo, P. E., Tang, S. H. E., Rentsendorj, A., Pfeifer, G. P., and Mann, J. R. (2000) Maternal-specific footprints at putative CTCF sites in the H19 imprinting control region give evidence for insulator function. *Current Biology* **10,** 607-610.

Tachibana, M., Sugimoto, K., Fukushima, T., and Shinkai, Y. (2001) Set domain-containing protein, G9a, is a novel lysine-preferring mammalian histone methyltransferase with hyperactivity and specific selectivity to lysines 9 and 27 of histone H3. *J. Biol. Chem.* **276,** 25309-25317.

Tachibana, M., Sugimoto, K., Nozaki, M., Ueda, J., Ohta, T., Ohki, M., Fukuda, M., Takeda, N., Niida, H., Kato, H., and Shinkai, Y. (2002) G9a histone methyltransferase plays a dominant role in euchromatic histone H3 lysine 9 methylation and is essential for early embryogenesis. *Genes Dev.* **16,** 1779-1791.

Tachibana, M., Matsumura, Y., Fukuda, M., Kimura, H., and Shinkai, Y. (2008) G9a/GLP complexes independently mediate H3K9 and DNA methylation to silence transcription. *EMBO Journal* **27,** 2681-2690.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001) Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. *Curr. Biol.* **11,** 1553-1558.

Tada, T., Tada, M., Hilton, K., Barton, S. C., Sado, T., Takagi, N., and Surani, M. A. (1998) Epigenotype switching of imprintable loci in embryonic germ cells. *Development Genes and Evolution* **207,** 551-561.

Tahiliani, M., Koh, K. P., Shen, Y. H., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009) Conversion of 5-Methylcytosine to 5-Hydroxymethylcytosine in Mammalian DNA by MLL Partner TET1. *Science* **324,** 930-935.

Takahashi, K. and Yamanaka, S. (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663-676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131,** 861-872.

Takai, D. and Jones, P. A. (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 3740-3745.

Tam, P. P. L. and Snow, M. H. L. (1981) Proliferation and Migration of Primordial Germ-Cells During Compensatory Growth in Mouse Embryos. *Journal of Embryology and Experimental Morphology* **64,** 133-147.

Tam, P. P. L., Zhou, S. X., and Tan, S. S. (1994) X-Chromosome Activity of the Mouse Primordial Germ-Cells Revealed by the Expression of An X-Linked Lacz Transgene. *Development* **120,** 2925-2932.

Tate, P. H. and Bird, A. P. (1993) Effects of DNA methylation on DNA-binding proteins and gene expression. *Curr. Opin. Genet. Dev.* **3,** 226-231.

Tatematsu, K., Yamazaki, T., and Ishikawa, F. (2000) MBD2-MBD3 complex binds to hemi-methylated DNA and forms a complex containing DNMT1 at the replication foci in late S phase. *Genes to Cells* **5,** 677-688.

Taylor, S. M. and Jones, P. A. (1982) Mechanism of Action of Eukaryotic DNA Methyltransferase - Use of 5-Azacytosine-Containing DNA. *Journal of Molecular Biology* **162,** 679-692.

Tazi, J. and Bird, A. (1990) Alternative Chromatin Structure at CpG Islands. *Cell* **60,** 909-920.

ten Berge, D., Koole, W., Fuerer, C., Fish, M., Eroglu, E., and Nusse, R. (2008) Wnt Signaling Mediates Self-Organization and Axis Formation in Embryoid Bodies. *Cell Stem Cell* **3,** 508-518.

Tesar, P. J., Chenoweth, J. G., Brook, F. A., Davies, T. J., Evans, E. P., Mack, D. L., Gardner, R. L., and Mckay, R. D. G. (2007) New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448,** 196-U10.

Ting, A. H., Jair, K. W., Suzuki, H., Yen, R. W., Baylin, S. B., and Schuebel, K. E. (2004) CpG island hypermethylation is maintained in human colorectal cancer cells after RNAi-mediated depletion of DNMT1. *Nat. Genet.* **36,** 582-584.

Ting, A. H., Jair, K. W., Suzuki, H., Yen, R. W., Baylin, S. B., and Schuebel, K. E. (2004) Mammalian DNA methyltransferase 1: inspiration for new directions. *Cell Cycle* **3,** 1024-1026.

Tomilin, N. V. (2008) Regulation of mammalian gene expression by retroelements and non-coding tandem repeats. *Bioessays* **30,** 338-348.

Tsukada, Y., Fang, J., Erdjument-Bromage, H., Warren, M. E., Borchers, C. H., Tempst, P., and Zhang, Y. (2006) Histone demethylation by a family of JmjC domain-containing proteins. *Nature* **439,** 811-816.

Tudor, M., Akbarian, S., Chen, R. Z., and Jaenisch, R. (2002) Transcriptional profiling of a mouse model for Rett syndrome reveals subtle transcriptional changes in the brain. *Proc. Natl. Acad. Sci. U.S.A.* **99,** 15536-15541.

Turek-Plewa, J. and Jagodzinski, P. P. (2005) The role of mammalian DNA methyltransferases in the regulation of gene expression. *Cell Mol. Biol. Lett.* **10,** 631-647.

Turner, B. M. (2000) Histone acetylation and an epigenetic code. *Bioessays* **22,** 836-845.

Tweedie, S., Charlton, J., Clark, V., and Bird, A. (1997) Methylation of genomes and genes at the invertebrate-vertebrate boundary. *Mol Cell Biol* **17,** 1469-1475.

Ueda, T., Abe, K., Miura, A., Yuzuriha, M., Zubair, M., Noguchi, M., Niwa, K., Kawase, Y., Kono, T., Matsuda, Y., Fujimoto, H., Shibata, H., Hayashizaki, Y., and Sasaki, H. (2000) The paternal methylation imprint of the mouse H19 locus is acquired in the gonocyte stage during foetal testis development. *Genes to Cells* **5,** 649-659.

Umar, A., Buermeyer, A. B., Simon, J. A., Thomas, D. C., Clark, A. B., Liskay, R. M., and Kunkel, T. A. (1996) Requirement for PCNA in DNA mismatch repair at a step preceding DNA resynthesis. *Cell* **87,** 65-73.

Umlauf, D., Goto, Y., Cao, R., Cerqueira, F., Wagschal, A., Zhang, Y., and Feil, R. (2004) Imprinting along the Kcnq1 domain on mouse chromosome 7 involves repressive histone methylation and recruitment of Polycomb group complexes. *Nature Genetics* **36,** 1296-1300.

Unoki, M., Nishidate, T., and Nakamura, Y. (2004) ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene* **23,** 7601-7610.

Vakoc, C. R., Sachdeva, M. M., Wang, H., and Blobel, G. A. (2006) Profile of histone lysine methylation across transcribed mammalian chromatin. *Mol. Cell Biol.* **26,** 9185-9195.

Vallender, T. W. and Lahn, B. T. (2006) Localized methylation in the key regulator gene endothelin-1 is associated with cell type-specific transcriptional silencing. *Febs Letters* **580,** 4560-4566.

Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol.* **3,** RESEARCH0034.

Vermeulen, M., Mulder, K. W., Denissov, S., Pijnappel, W. W., van Schaik, F. M., Varier, R. A., Baltissen, M. P., Stunnenberg, H. G., Mann, M., and Timmers, H. T. (2007) Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* **131,** 58-69.

Vertino, P. M., Sekowski, J. A., Coll, J. M., Applegren, N., Han, S., Hickey, R. J., and Malkas, L. H. (2002) DNMT1 is a component of a multiprotein DNA replication complex. *Cell Cycle* **1,** 416-423.

Vilkaitis, G., Suetake, I., Klimasauskas, S., and Tajima, S. (2005) Processive methylation of hemimethylated CpG sites by mouse Dnmt1 DNA methyltransferase. *J Biol Chem.* **280,** 64-72.

Vincent, S. D., Dunn, N. R., Sciammas, R., Shapiro-Shalef, M., Davis, M. M., Calame, K., Bikoff, E. K., and Robertson, E. J. (2005) The zinc finger transcriptional repressor Blimp1/Prdm1 is dispensable for early axis formation but is required for specification of primordial germ cells in the mouse. *Development* **132,** 1315-1325.

Vire, E., Brenner, C., Deplus, R., Blanchon, L., Fraga, M., Didelot, C., Morey, L., Van Eynde, A., Bernard, D., Vanderwinden, J. M., Bollen, M., Esteller, M., Di Croce, L., de Launoit, Y., and Fuks, F. (2006) The Polycomb group protein EZH2 directly controls DNA methylation. *Nature* **439,** 871-874.

Waddington, C.H. (1942) (1977). "The epigenotype". *Endeavour* **1**: 18–20.

Waddington, C.H. (1957) *The strategy of the genes*.

Walsh, C. P., Chaillet, J. R., and Bestor, T. H. (1998) Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nature Genetics* **20,** 116-117.

Walter, W., Clynes, D., Tang, Y., Marmorstein, R., Mellor, J., and Berger, S. L. (2008) 14-3-3 interaction with histone H3 involves a dual modification pattern of phosphoacetylation. *Mol. Cell Biol.* **28,** 2840-2849.

Wang, H., Zhai, L., Xu, J., Joo, H. Y., Jackson, S., Erdjument-Bromage, H., Tempst, P., Xiong, Y., and Zhang, Y. (2006) Histone H3 and H4 ubiquitylation by the CUL4-DDB-ROC1 ubiquitin ligase facilitates cellular response to DNA damage. *Mol. Cell* **22,** 383-394.

Wang, P. J., McCarrey, J. R., Yang, F., and Page, D. C. (2001) An abundance of X-linked genes expressed in spermatogonia. *Nat. Genet.* **27,** 422-426.

Watanabe, D., Suetake, I., Tada, T., and Tajima, S. (2002) Stage- and cell-specific expression of Dnmt3a and Dnmt3b during embryogenesis. *Mechanisms of Development* **118,** 187-190.

Watt, F. and Molloy, P. L. (1988) Cytosine Methylation Prevents Binding to DNA of A Hela-Cell Transcription Factor Required for Optimal Expression of the Adenovirus Major Late Promoter. *Genes Dev* **2,** 1136-1143.

Weber, M. and Schubeler, D. (2007) Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Current Opinion in Cell Biology* **19,** 273-280.

Weber, M., Hellmann, I., Stadler, M. B., Ramos, L., Paabo, S., Rebhan, M., and Schubeler, D. (2007) Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nat. Genet.* **39,** 457-466.

Webster, K. E., O'Bryan, M. K., Fletcher, S., Crewther, P. E., Aapola, U., Craig, J., Harrison, D. K., Aung, H., Phutikanit, N., Lyle, R., Meachem, S. J., Antonarakis, S. E., de Krester, D. M., Hedger, M. P., Peterson, P., Carroll, B. J., and Scott, H. S. (2005) Meiotic and epigenetic defects in Dnmt3L-knockout mouse spermatogenesis. *Proc. Natl. Acad. Sci. U.S.A.* **102,** 4068-4073.

Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A., and Feinberg, A. P. (2009) Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nat. Genet.* **41,** 246-250.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B. E., and Jaenisch, R. (2007) In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature*.

Wernig, M., Zhao, J. P., Pruszak, J., Hedlund, E., Fu, D. D., Soldner, F., Broccoli, V., Constantine-Paton, M., Isacson, O., and Jaenisch, R. (2008) Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. *Proc. Natl. Acad. Sci. U.S.A.* **105,** 5856-5861.

Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A., and Gough, N. M. (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336,** 684-687.

Wilmut, I., Schnieke, A. E., McWhir, J., Kind, A. J., and Campbell, K. H. S. (1997) Viable offspring derived from fetal and adult mammalian cells. *Nature* **385,** 810-813.

Wolffe, A. P. and Matzke, M. A. (1999) Epigenetics: regulation through repression. *Science* **286,** 481-486.

Woodcock, C. L., Frado, L. L., and Rattner, J. B. (1984) The higher-order structure of chromatin: evidence for a helical ribbon arrangement. *J. Cell Biol.* **99,** 42-52.

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., and Romano, L. A. (2003) The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* **20,** 1377-1419.

Wu, J.C. and Santi, D.V. (1985) On the mechanism and methyltransferases. *Prog Clin Biol Res*, **198**, 119-129.

Wu, J.C. and Santi, D. V. (1987) Kinetic and Catalytic Mechanism of Hhai Methyltransferase. *J Biol Chem.* **262,** 4778-4786.

Wu, J., Issa, J. P., Herman, J., Bassett, D. E., Jr., Nelkin, B. D., and Baylin, S. B. (1993) Expression of an exogenous eukaryotic DNA methyltransferase gene induces transformation of NIH 3T3 cells. *Proc. Natl. Acad. Sci. U. S. A* **90,** 8891-8895.

Wu, Q., Chen, X., Zhang, J. Q., Loh, Y. H., Low, T. Y., Zhang, W. W., Zhang, W. S., Sze, S. K., Lim, B., and Ng, H. H. (2006) Sall4 interacts with nanog and co-occupies nanog genomic sites in embryonic stem cells. *J Biol Chem.* **281,** 24090-24094.

Wyatt, G.R**.** (1950) Occurrence of 5-methylcytosine in nucleic acids. *Nature*, **166**, 237-238.
Xi, S., Geiman, T. M., Briones, V., Guang, T. Y., Xu, H., and Muegge, K. (2009) Lsh participates in DNA methylation and silencing of stem cell genes. *Stem Cells* **27,** 2691-2702.

Xi, S. C., Zhu, H. M., Xu, H., Schmidtmann, A., Geiman, T. M., and Muegge, K. (2007) Lsh controls Hox gene silencing during development. *Proc. Natl. Acad. Sci. U.S.A.* **104,** 14366-14371.

Xiao, L., Wang, Y., Zhou, Y., Sun, Y., Sun, W., Wang, L., Zhou, C., Zhou, J., and Zhang, J. (2009) Identification of a novel human cancer/testis gene MAEL that is regulated by DNA methylation. *Mol. Biol. Rep.*

Xie, S., Wang, Z., Okano, M., Nogami, M., Li, Y., He, W. W., Okumura, K., and Li, E. (1999) Cloning, expression and chromosome locations of the human DNMT3 gene family. *Gene* **236,** 87-95.

Yang, J., Chai, L., Fowles, T. C., Alipio, Z., Xu, D., Fink, L. M., Ward, D. C., and Ma, Y. (2008) Genome-wide analysis reveals Sall4 to be a major regulator of pluripotency in murine-embryonic stem cells. *Proc. Natl. Acad. Sci. U. S A* **105,** 19756-19761.

Yasunaga, M., Tada, S., Nishikawa, S. T., Nakano, Y., Okada, M., Jakt, L. M., Nishikawa, S., Chiba, T., Era, T., and Nishikawa, S. I. (2005) Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nature Biotechnology* **23,** 1542-1550.

Ying, Q. L. and Smith, A. G. (2003) Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365,** 327-341.

Ying, Q. L., Nichols, J., Chambers, I., and Smith, A. (2003) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115,** 281-292.

Ying, Q. L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008) The ground state of embryonic stem cell self-renewal. *Nature* **453,** 519-523.

Yoder, J. A., Soman, N. S., Verdine, G. L., and Bestor, T. H. (1997) DNA (cytosine-5)-methyltransferases in mouse cells and tissues. Studies with a mechanism-based probe. *J. Mol. Biol.* **270,** 385-395.

Yoder, J. A. and Bestor, T. H. (1998) A candidate mammalian DNA methyltransferase related to pmt1p of fission yeast. *Hum. Mol. Genet.* **7,** 279-284.

Yokochi, T., Poduch, K., Ryba, T., Lu, J. J., Hiratani, I., Tachibana, M., Shinkai, Y., and Gilbert, D. M. (2009) G9a selectively represses a class of late-replicating genes at the nuclear periphery. *Proc. Natl. Acad. Sci. U.S.A.* **106,** 19363-19368.

Yoon, H. G., Chan, D. W., Reynolds, A. B., Qin, J., and Wong, J. M. (2003) N-CoR mediates DNA methylation-dependent repression through a methyl CpG binding protein Kaiso. *Mol. Cell* **12,** 723-734.

Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S., Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, I. I., and Thomson, J. A. (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318,** 1917-1920.

Yu, W. Q., Gius, D., Onyango, P., Muldoon-Jacobs, K., Karp, J., Feinberg, A. P., and Cui, H. M. (2008) Epigenetic silencing of tumour suppressor gene p15 by its antisense RNA. *Nature* **451,** 202-U10.

Yuan, P., Han, J. Y., Guo, G. J., Orlov, Y. L., Huss, M., Loh, Y. H., Yaw, L. P., Robson, P., Lim, B., and Ng, H. H. (2009) Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes Dev* **23,** 2507-2520.

Zhang, J., Xu, F., Hashimshony, T., Keshet, I., and Cedar, H. (2002) Establishment of transcriptional competence in early and late S phase. *Nature* **420,** 198-202.

Zhang, J. Q., Tam, W. L., Tong, G. Q., Wu, Q., Chan, H. Y., Soh, B. S., Lou, Y. F., Yang, J. C., Ma, Y. P., Chai, L., Ng, H. H., Lufkin, T., Robson, P., and Lim, B. (2006) Sall4 modulates embryonic stem cell pluripotency and early embryonic development by the transcriptional regulation of Pou5f1. *Nature Cell Biology* **8,** 1114-U125.

Zhang, L. P., Stroud, J. C., Walter, C. A., Adrian, G. S., and McCarrey, J. R. (1998) A gene-specific promoter in transgenic mice directs testis-specific demethylation prior to transcriptional activation in vivo. *Biology of Reproduction* **59,** 284-292.

Zhang, X. and Verdine, G. L. (1996) Mammalian DNA cytosine-5 methyltransferase interacts with p23 protein. *FEBS Lett.* **392,** 179-183.

Zhang, Y., Ng, H. H., Erdjument-Bromage, H., Tempst, P., Bird, A., and Reinberg, D. (1999) Analysis of the NuRD subunits reveals a histone deacetylase core complex and a connection with DNA methylation. *Genes Dev* **13,** 1924-1935.

Zhang, Y. and Reinberg, D. (2001) Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails. *Genes Dev.* **15,** 2343-2360.

Zhao, R. and Daley, G. Q. (2008) From Fibroblasts to iPS Cells: Induced Pluripotency by Defined Factors. *Journal of Cellular Biochemistry* **105,** 949-955.

Zhao, X. Y., Ueba, T., Christie, B. R., Barkho, B., McConnell, M. J., Nakashima, K., Lein, E. S., Eadie, B. D., Willhoite, A. R., Muotri, A. R., Summers, R. G., Chun, J., Lee, K. F., and Gage, F. H. (2003) Mice lacking methyl-CpG binding protein 1 have deficits in adult neurogenesis and hippocampal function. *Proc. Natl. Acad. Sci. U.S.A.* **100,** 6777-6782.

Zhu, B., Zheng, Y., Pham, A. D., Mandal, S. S., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2005) Monoubiquitination of human histone H2B: the factors involved and their roles in HOX gene regulation. *Mol. Cell* **20,** 601-611.

Zhu, H. M., Geiman, T. M., Xi, S. C., Jiang, Q., Schmidtmann, A., Chen, T. P., Li, E., and Muegge, K. (2006) Lsh is involved in de novo methylation of DNA. *EMBO Journal* **25,** 335-345.

Zhu, W. G., Srinivasan, K., Dai, Z. Y., Duan, W. R., Druhan, L. J., Ding, H. M., Yee, L., Villalona-Calero, M. A., Plass, C., and Otterson, G. A. (2003) Methylation of adjacent CpG sites affects Sp1/Sp3 binding and activity in the p21(Cip1) promoter. *Mol Cell Biol* **23,** 4056-4065.

Zilberman, D. and Henikoff, S. (2007) Genome-wide analysis of DNA methylation patterns. *Development* **134,** 3959-3965.

Zilberman, D., Coleman-Derr, D., Ballinger, T., and Henikoff, S. (2008) Histone H2A.Z and DNA methylation are mutually antagonistic chromatin marks. *Nature* **456,** 125-U14.

Zupkovitz, G., Tischler, J., Posch, M., Sadzak, I., Ramsauer, K., Egger, G., Grausenburger, R., Schweifer, N., Chiocca, S., Decker, T., and Seiser, C. (2006) Negative and positive regulation of gene expression by mouse histone deacetylase 1. *Mol. Cell Biol.* **26,** 7913-7928.

Publications arising

Throughout my PhD I have worked on several projects aside from that presented here, particularly on the roles and functions of Dnmt1 and Kaiso during *Xenopus* development. Shown overleaf are the publications that have resulted from these studies.

An additional chapter to this thesis entitled “The non-catalytic roles of Dnmt1 in mammals” (~30 pages) is available as an electronic supplement or upon request.

jhackett@hgu.mrc.ac.uk

Development 136, 723-727 (2009) doi:10.1242/dev.025577

The interaction of xKaiso with xTcf3: a revised model for integration of epigenetic and Wnt signalling pathways

Alexey Ruzov[1], Jamie A. Hackett[1], Anna Prokhortchouk[2], James P. Reddington[1], Monika J. Madej[1], Donncha S. Dunican[1], Egor Prokhortchouk[2], Sari Pennings[3] and Richard R. Meehan[1,*]

We demonstrate that a direct interaction between the methyl-CpG-dependent transcription repressor Kaiso and xTcf3, a transducer of the Wnt signalling pathway, results in their mutual disengagement from their respective DNA-binding sites. Thus, the transcription functions of xTcf3 can be inhibited by overexpression of Kaiso in cell lines and *Xenopus* embryos. The interaction of Kaiso with xTcf3 is highly conserved and is dependent on its zinc-finger domains (ZF1-3) and the corresponding HMG DNA-binding domain of TCF3/4 factors. Our data rule out a model suggesting that xKaiso is a direct repressor of Wnt signalling target genes in early *Xenopus* development via binding to promoter-proximal CTGCNA sequences as part of a xTcf3 repressor complex. Instead, we propose that mutual inhibition by Kaiso/TCF3 of their DNA-binding functions may be important in developmental or cancer contexts and acts as a regulatory node that integrates epigenetic and Wnt signalling pathways.

KEY WORDS: Cancer, DNA methylation, Kaiso, Siamois, TCF3, Chromatin

INTRODUCTION

The BTB/POZ transcriptional factor xKaiso is a bimodal DNA-binding protein that is reported to specifically bind methyl-CpGs, or a CTGCNA consensus DNA sequence (Ruzov et al., 2004; Park et al., 2005). Our previous work has established the essential and global role of xKaiso in regulating the timing of zygotic gene activation at the mid-blastula transition (MBT) (Ruzov et al., 2004). Other work proposes a model in which xKaiso specifically binds CTGCNA sequences present in the promoter region of Siamois (and also xWnt11) and interacts with the Wnt effector molecule xTcf3 to promote its stable repression (Kim et al., 2004; Park et al., 2005).

In recent work that is not in keeping with the latter model, we demonstrated that the CTGCNA motifs derived from the promoters of Siamois and xWnt11 are not sequence specific xKaiso-binding sites and these genes are not mis-expressed in xKaiso morphants (Ruzov et al., 2009). Although our loss-of-function experiments did not identify a role for xKaiso in regulating Wnt target genes, two observations suggest a potential role for it in canonical Wnt signalling: the xKaiso protein can be co-immunoprecipitated with xTcf3, and over-expression of xKaiso can suppress axis-duplication that is induced by over-expression of β-catenin in the ventral regions of a four-cell embryo (Park et al., 2005). We wished to determine the molecular basis of the Kaiso/Tcf3 interaction as it has profound implications for the intersection of two important regulatory pathways in amphibian development: regulation of transcriptional silencing in pre-MBT *Xenopus* embryos and in Wnt signalling pathways (Heasman, 2006). We find that the interaction surfaces for both proteins correspond to their previously identified DNA-binding domains. Our data squarely rules out a model for xKaiso repression through stabilisation of xTcf3 binding to DNA. Instead, our analysis suggests that the xKaiso and xTcf3/4 interaction results in their mutual delocalisation from chromatin, a prediction that we demonstrate at the cellular and the DNA levels.

[1]Human Genetics Unit, MRC, Western General Hospital, Edinburgh EH4 2XU, UK. [2]Center 'Bioengineering', 60-let Oktyabrya 7-1, Moscow, 117312, Russian Federation. [3]Queen's Medical Research Institute, University of Edinburgh, Edinburgh EH16 4TJ, UK.

*Author for correspondence (e-mail: r.meehan@hgu.mrc.ac.uk)

Accepted 9 December 2008

MATERIALS AND METHODS

Reporter assays and expression constructs

The xKaiso expression constructs (XKaiso/pCS2+MT and HA-tagged) (Kim et al., 2004) were provided by Pierre McCrea. The dKaiso expression construct was from (Ruzov et al., 2004). The Myc-xTcf3, xTcf3dn and β-catenin expression vectors were provided by Randall Moon. The VP16 fusions with the ZF1-3 region of xKaiso (amino acids 447-635) and the HMG domain of xTcf3 were from Ruzov et al. (Ruzov et al., 2009). All reporter assays using SuperTOP/FOP, Tex19 and Siamois luciferase reporters were performed as described previously (Houston et al., 2002; Ruzov et al., 2009).

Embryos and microinjections

Embryos were manipulated as described previously (Houston et al., 2002; Ruzov et al., 2009). At the two-cell stage, the embryos were injected into the animal half with 200-750 pg of sense capped RNA (c-myc-xKaiso mRNA) synthesized in vitro (T3/T7 Cap-Scribe kit, Boehringer).

GST pull-down assays, immunoprecipitation and EMSA

The Kaiso GST fusions were from Ruzov et al. (Ruzov et al., 2009). The TCF4 constructs were provided by Vladimir Korinek. A coupled transcription/translation kit (Promega) was used for in vitro translation/labelling with ^{35}S-Met. GST pull-down assays and immunoprecipitations were performed according to standard protocols in the presence of a 10- to 100-fold excess of recombinant full-length xKaiso or GST where indicated. Samples were visualized by phosphoimaging. EMSA was performed as described previously (Ruzov et al., 2009).

Chromatin immunoprecipitation (ChIP) assay

The ChIP assay was performed as described previously (Ruzov et al., 2009) using myc tagged xTcf3.

Immunostaining

Immunostaining was performed according to standard techniques using *P53*$^{-/-}$ (*Trp53*$^{-/-}$) mouse embryonic fibroblasts. Cells were analysed 24 hours after transfection. Mouse monoclonal anti-T7 tag (Novagen), anti-HA-tag (Sigma), rabbit polyclonal anti-myc (Upstate) and Alexa secondary antibodies were used.

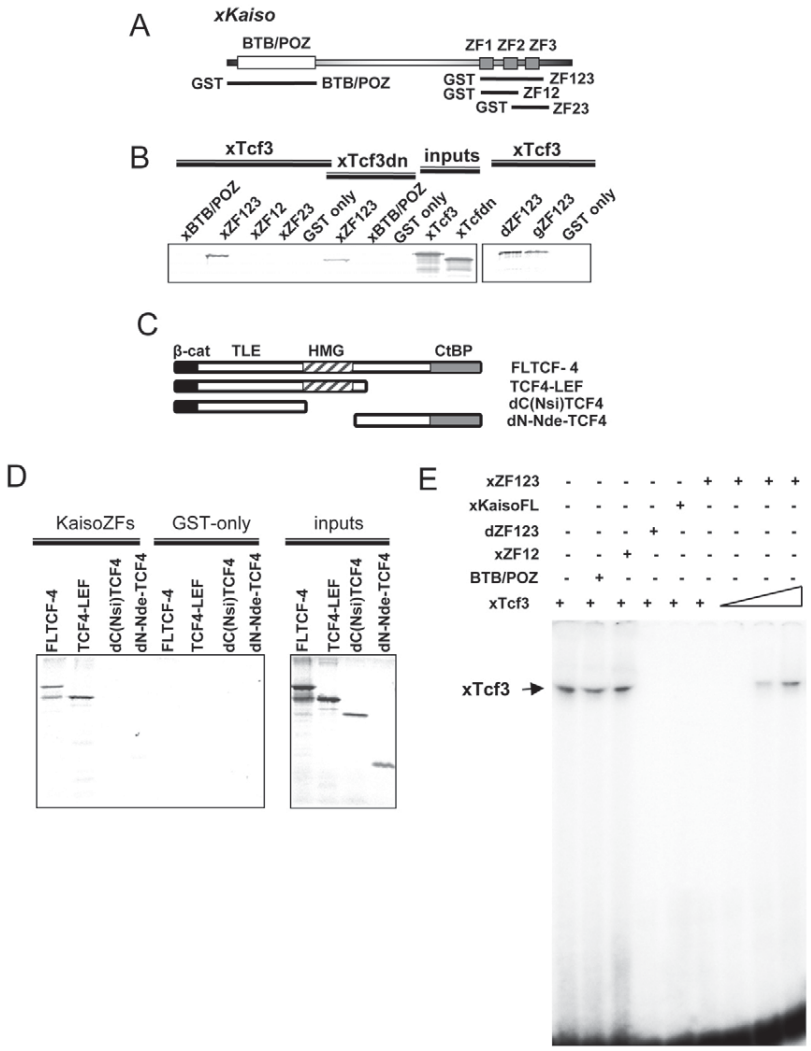


Fig. 1. Kaiso ZF1-3 interacts directly with the HMG domain of TCF factors. (**A**) Kaiso deletion constructs used in pull-down experiments. (**B**) GST pull-down using in vitro transcribed xTcf3 or xTcf3dn with Kaiso recombinant proteins indicated in A. GST was used as a control. (**C**) TCF4 constructs used for in vitro transcription in D. β-cat, β-catenin interaction domain; TLE, TLE/Groucho binding domain; CtBP, CtBP-binding sites; HMG, DNA-binding domain. (**D**) GST pull-down using TCF4 constructs and xKaiso GST-ZF domain (KaisoZFs). GST was used as a control; 1/10 of inputs (xTcf3 and xTcf3dn) are shown. (**E**) EMSA assay using in vitro translated xTcf3 and its DNA-binding site (TCF bs-oligo) in the absence and presence of GST fusions with ZF1-3 of xKaiso, full-length xKaiso or ZF1-3 of dKaiso. xZF2-3 and BTB/POZ fusions, which do not interact with xTCF3 were used as controls. Note absence of the xTcf3/DNA complex (arrow) in the presence of the Kaiso fusions. Increasing the xTcf3 concentration in the presence of xZF1-3 restores the binding of xTcf3 to DNA, supporting the hypothesis that sequestration is the mechanism of inhibition of xTcf3 binding to DNA by Kaiso.

Real-time RT-PCR

Quantitative real-time RT-PCR of Siamois was evaluated as described previously (Houston et al., 2002).

RESULTS AND DISCUSSION

Kaiso directly interacts with the HMG domain of TCF factors via ZF1-3

The TCF family proteins have a multi-domain organisation with a central HMG box DNA-binding region recognizing the sequence A/TA/TCAAA; an N terminus containing the β-catenin-interacting domain adjacent to a Groucho-binding region; and a C-terminal CtBP1 interaction domain (Roose et al., 1998). We demonstrate using recombinant proteins that the DNA-binding, zinc-finger domain of xKaiso (xZF1-3) is sufficient for direct interaction with full-length xTcf3 and dominant-negative xTcf3 (xTcf3dn), which lacks the β-catenin-interacting region (Fig. 1A,B). The ability of Kaiso to bind xTcf3 is conserved, as comparable zinc-finger regions from zebrafish Kaiso (dZF1-3) and chicken Kaiso (gZF1-3) can also interact with xTcf3 (Fig. 1B). The same pattern of interaction is seen between mouse TCF4 and xZF1-3. Deletion analysis suggests that the interaction occurs through the HMG domain of TCF4 (Fig. 1C,D). The experiment was performed in the presence of high concentrations of ethidium bromide to exclude the possibility that the interaction was mediated by non-specific binding to DNA. This suggests that the interaction between Kaiso and TCF3/4 is mutually exclusive of their binding to DNA, and that inhibition of β-catenin activation by Kaiso is not through competitive binding of a shared interaction domain on xTcf3 and TCF4. We tested the first possibility by performing an EMSA with xTcf3 and its target DNA binding site (ATCAAA) in the presence of GST-Kaiso fusions. Binding of xTcf3 to its target sequence was abolished in the presence of the full-length xKaiso, GST-xZF1-3 and GST-dZF1-3. GST alone or GST-xZF1-2 had no effect (Fig. 1E). Increasing the amount of xTcf3 overcomes the inhibitory effect of GST-xZF1-3. Moreover, the interaction between xTcf3 and β-catenin in vitro is not compromised by xZF1-3 (Fig. 2A). We also examined the cellular location of myc-tagged xTcf3 in the absence and presence of Kaiso. xTcf3 localisation in mouse cells is nuclear, with about 40% of cells with a homogenous, as opposed to a speckled, pattern (60%) (Fig. 2B); the latter is reminiscent of the discrete foci observed for endogenous mouse TCF4 (Valenta et al., 2006). xKaiso in mouse fibroblasts exhibits homogenous nuclear staining (Fig. 2C). In the presence of either xKaiso or dKaiso, the pattern of xTcf3 staining was uniformly homogenous, suggesting they alter its nuclear sublocalisation (Fig. 2D).

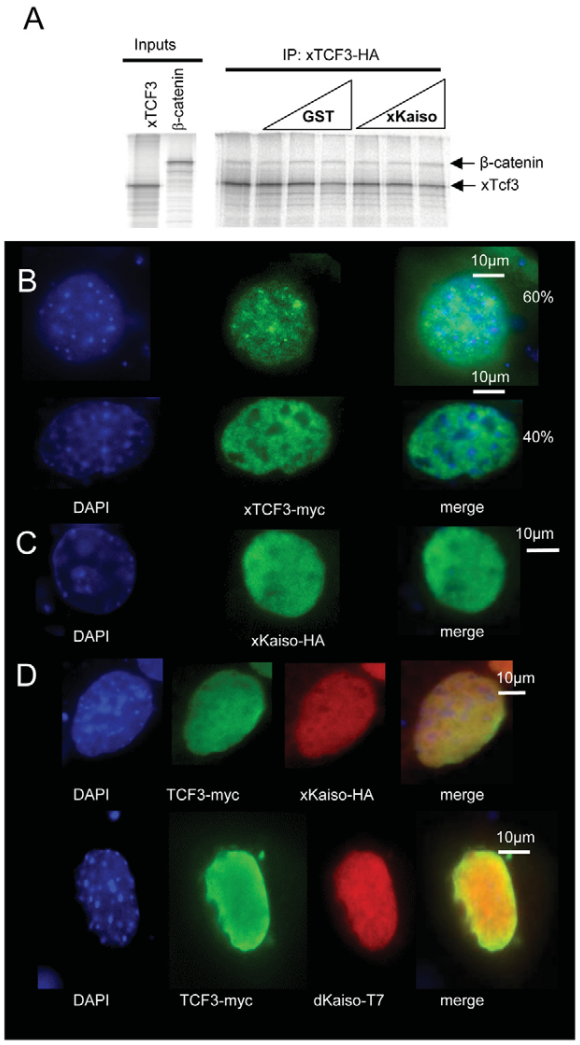


Fig. 2. Kaiso does not affect β-catenin interaction with xTcf3 but can alter the nuclear localisation of TCF3 in mouse fibroblasts. (**A**) In vitro translated ^{35}S-methione labelled β-catenin and HA-tagged xTcf3 were incubated in the presence of an excess of xKaiso full-length (xKaisoFL) protein or GST. xTcf3 was immunoprecipitated with an anti-HA antibody. β-Catenin and xTcf3 proteins are indicated by arrows. 1/10 of inputs are shown. The interaction between β-catenin and xTcf3 is not impaired in the presence of GST-xKaiso or GST. (B,C) Localisation of myc-xTcf3 upon transient transfection into mouse MEFs in the absence and presence of Ha-xKaiso and T7tagged-dKaiso. (**B**) By itself, myc-xTcf3 either exhibits staining at nuclear foci (60% of cells) or homogenous nuclear staining (40%). (**C**) xKaiso by itself exhibits homogenous nuclear staining in 100% of cells. (**D**) In the presence of xKaiso or dKaiso the myc-xTcf3 protein exhibits only homogenous nuclear staining (100% of cells with double staining).

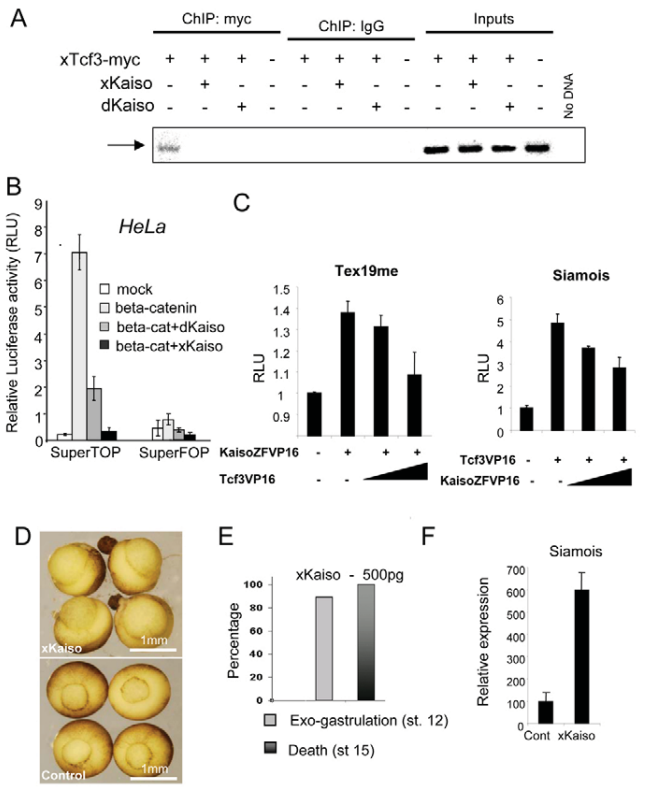


Fig. 3. Kaiso prevents xTcf3 localising to its target promoter and its subsequent activation or repression functions. (**A**) ChIP assay showing that, after transient transfection into A6 cells, myc-xTcf3 can be located at the *Siamois* promoter. Co-expression of xKaiso or dKaiso prevents xTcf3 binding at the promoter. Inputs are shown on the right and the myc-Tcf3 ChIP on the left with a non-specific antibody (Ig) control. (**B**) Overexpression of both dKaiso and xKaiso represses β-catenin-dependent transcription activity in transient transfection assays in HeLa cells. SuperTopflash (SuperTop) was used as a Wnt reporter and SuperFopflash (SuperFop) with mutated TCF3 sites as a control. (**C**) Expression from a methylated luciferase reporter can be enhanced by co-transfection with an xKaisoZFVP16 (KaisoZFVP16) expression plasmid. Targeting to the methylated reporter is provided by the Kaiso-ZFs and transcription activation function by VP16. Increasing amounts of xTcf3HMGVP16 (Tcf3VP16) reduce the activation potential of xKaisoZFVP16. In a reciprocal experiment, activation a Siamois luciferase reporter by xTcf3HMGVP16 is inhibited by the presence of xKaisoZFVP16. (**D**,**E**) Overexpression of myc-xKaiso in *Xenopus* embryos results in the formation of exogastrulae compared with controls at stage 12. Graph shows that by stage 12, over 80% of the embryos (n=55) exhibit exogastrulae. (**F**) Real-time RT-PCR analysis of Siamois expression in stage 10 control and embryos injected with myc-xKaiso mRNA (1 ng). Relative to histone, H4 Siamois expression is activated in the myc-xKaiso-injected embryos.

xKaiso displaces xTcf3 from its target promoters

The interaction data suggests that overexpression of xKaiso can displace xTcf3 from its genomic binding sites. Using the chromatin immunoprecipitation (ChIP) technique we could localise myc-xTcf3 to the *Siamois* promoter in A6 cells (Fig. 3A). However, myc-xTcf3 was delocalised in the presence of either Ha-xKaiso or T7tag-dKaiso (Fig. 3A). In contrast to the model of Park et al. (Park et al., 2005), this suggests that Kaiso can disengage xTcf3 from its target genes and alter their expression state. Under the same conditions, we could not localize dKaiso or xKaiso alone to the Siamois promoter; this is not surprising in view of the absence of high-affinity CTGCNA or methyl-CpG binding sites in it (Ruzov et al., 2009).

The SuperTopflash Wnt reporter plasmid contains the luciferase gene under the control of multiple TCF3/4-binding sites, which is activated by the addition of exogenous β-catenin and endogenous TCF3/4 in HeLa cells (Veeman et al., 2003) (Fig. 3B). This

activation was inhibited between four- to sevenfold by co-expression of either xKaiso or dKaiso (Fig. 3B). The transcriptional activity of a control reporter with mutated TCF3-binding sites is unresponsive to β-catenin induction and unaffected by the presence of either Kaiso (Fig. 3B). An obvious explanation of these results is that the Kaiso interaction with the HMG domain of endogenous TCF3/4 blocks its ability to bind DNA and inhibits β-catenin-dependent activation of transcription (Fig. 3B). Conversely, overexpression of xTcf3 interferes with transcriptional activation of a methylated reporter template by an xKaisoZFVP16 fusion (Fig. 3C) in a dose-dependent manner suggesting the Kaiso/TCF3 interaction results in a mutual disengagement of these factors from chromatin-binding sites. This model can account for the inhibition of dorsal axis formation when Kaiso mRNA is co-injected with β-catenin mRNA (Park et al., 2005) as excess xKaiso will displace the xTcf3/β-catenin complex and inhibit axis duplication. Similarly, xKaisoZFVP16 directly inhibits xTcf3HMGVP16 activation of a Siamois reporter in a dose-dependent manner (Fig. 3C).

Loss-of-function studies in *Xenopus* suggest the primary role of xTcf3 is as a repressor of organiser genes, such as Siamois, throughout the early embryo (Houston et al., 2002). This repression is inactivated on the dorsal side by a maternally encoded Wnt signalling pathway. Based on the above results, we predicted that overexpression of xKaiso in developing *Xenopus* embryos would mimic certain aspects of xTcf3 depletion such as ectopic Siamois expression (Houston et al., 2002). To test this, we injected xKaiso mRNA into two-cell embryos and allowed them to develop until gastrulation (stage 10.5-11) (Fig. 3D). In comparison with controls, this results in an exogastrulae phenotype that is distinct from the developmental delay phenotype observed in xKMO morphants, but is similar to the phenotype of embryos that are maternally depleted of xTcf3 (Fig. 3D,E) (Houston et al., 2002; Ruzov et al., 2009). The normalised (relative to histone H4 or GAPDH levels) expression levels of Siamois are increased in the xKaiso mRNA-injected embryos (Fig. 3F). This observation suggests that high levels of xKaiso can interfere with xTcf3 function, resulting in a relative relief of xTcf3/Groucho-mediated repression of Siamois in gastrulating embryos.

These new results call for a revision of the model connecting the xKaiso repressor function and the xTCF3 repression/activation of Wnt-target genes (Kim et al., 2004; Park et al., 2005). The mode of interaction between xKaiso and xTCF3 mutually prevents binding to their cognate DNA sites, which can potentially inhibit the two pathways in which these transcription factors are major participants (Heasman, 2006). However, the recent observation that neither xWnt11 or Siamois expression is altered in xKMO morphants suggests that there is no intersection between these pathways during gastrulation (Ruzov et al., 2009).

Is this model operative in other biological contexts?

The role of Kaiso expressed in adult somatic tissues is not clear as there is no obvious mis-expression of normally silent genes in Kaiso-null mice (Prokhortchouk et al., 2006). In cancer cells, gene expression patterns are highly disturbed and this is mirrored by alterations in the level and genomic distribution of DNA methylation, as well as histone modifications (Ohm et al., 2007). Interestingly, Kaiso levels are increased in colon cancer cells and enhance polyp formation in Min mice ($Apc^{Min/+}$) (Prokhortchouk et al., 2006). This is juxtaposed by positional differences in TCF variant expression in colon cancer (Clevers, 2006). In this case,

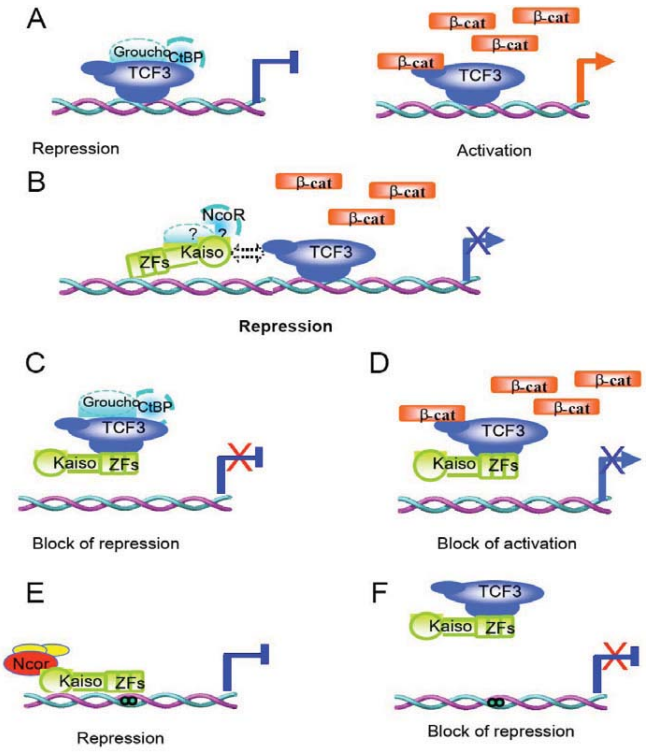


Fig. 4. Models for mutual interference of DNA-binding functions by Kaiso and TCF3. (**A**) A schematic representation of β-catenin-dependent activation of TCF3 target genes. β-Catenin displaces the Groucho/CtBP complex from TCF3 leading to gene activation. (**B**) A model for Kaiso-mediated repression of Wnt target genes that was suggested by Park et al. (Park et al., 2005). Here, it was proposed that repression of the Siamois gene occurs by xKaiso binding to CTGCNA sequences in its promoter and the possible interaction pf Kaiso with xTcf3. Our results suggest this model is not operative (Ruzov et al., 2009). (**C**,**D**) Our model of non-DNA-dependent xKaiso displacement of xTcf3 from target genes proposes that Kaiso overexpression may result in either a block of repression by xTcf3/Groucho or a block of β-catenin/xTcf3-dependent activation. (**E**,**F**) TCF3 can potentially displace Kaiso from its methylated (filed circles) DNA-binding sites and interfere with its repression function.

there is a possible intersection of Kaiso with Wnt signalling pathways, where overexpression of Kaiso could attenuate constitutive Wnt signalling, while at the same time promote cancer progression through silencing of de novo methylated tumour suppressor genes. By the same token, there is a potential for TCF3/4 to displace Kaiso from its cancer target genes and promote their re-expression, which may be coincident with alterations in the tumour environment, such as the transition to metastases phenotype (Fig. 4E). There are two related Kaiso-like proteins, Zbtb4 and Zbtb38, that may also mediate an intersection between epigenetic and cellular signalling pathways in cancer via protein-protein interactions (Filion et al., 2006; Weber et al., 2008). Zbt4 has been shown to inhibit MIZ1 regulation of p21CIP1 expression possibly by displacing this transcription factor from the p21CIP1 promoter (Weber et al., 2008). This suggests that the distributions of these Kaiso-like proteins and their relative abundance within nuclear sub-compartments has the potential to determine their transcriptional output of many signalling pathways.

We thank Vladimir Korinek, Pierre McCrea and Randall Moon for reagents, and Nicola Grey and Bill Richardson for *Xenopus* embryos. Work in R.M.'s laboratory (R.M., D.D., A.R. and J.H.) is supported by the MRC. E.P. and A.P. were supported by Russian Academy of Sciences program grant 'Molecular and Cellular Biology', RFBR 06-04-49216-a and CRDF RB-2808. Deposited in PMC for release after 6 months.

References

Clevers, H. (2006). Wnt/beta-catenin signaling in development and disease. *Cell* **127**, 469-480.

Filion, G. J., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E. and Defossez, P. A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol. Cell. Biol.* **26**, 169-181.

Heasman, J. (2006). Patterning the early Xenopus embryo. *Development* **133**, 1205-1217.

Houston, D. W., Kofron, M., Resnik, E., Langland, R., Destree, O., Wylie, C. and Heasman, J. (2002). Repression of organizer genes in dorsal and ventral Xenopus cells mediated by maternal XTcf3. *Development* **129**, 4015-4025.

Kim, S. W., Park, J. I., Spring, C. M., Sater, A. K., Ji, H., Otchere, A. A., Daniel, J. M. and McCrea, P. D. (2004). Non-canonical Wnt signals are modulated by the Kaiso transcriptional repressor and p120-catenin. *Nat. Cell Biol.* **6**, 1212-1220.

Ohm, J. E., McGarvey, K. M., Yu, X., Cheng, L., Schuebel, K. E., Cope, L., Mohammad, H. P., Chen, W., Daniel, V. C., Yu, W. et al. (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat. Genet.* **39**, 237-242.

Park, J. I., Kim, S. W., Lyons, J. P., Ji, H., Nguyen, T. T., Cho, K., Barton, M. C., Deroo, T., Vleminckx, K., Moon, R. T. and McCrea, P. D. (2005). Kaiso/p120-catenin and TCF/beta-catenin complexes coordinately regulate canonical Wnt gene targets. *Dev. Cell* **8**, 843-854.

Prokhortchouk, A., Sansom, O., Selfridge, J., Caballero, I. M., Salozhin, S., Aithozhina, D., Cerchietti, L., Meng, F. G., Augenlicht, L. H., Mariadason, J. M. et al. (2006). Kaiso-deficient mice show resistance to intestinal cancer. *Mol. Cell. Biol.* **26**, 199-208.

Roose, J., Molenaar, M., Peterson, J., Hurenkamp, J., Brantjes, H., Moerer, P., van de Wetering, M., Destree, O. and Clevers, H. (1998). The Xenopus Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. *Nature* **395**, 608-612.

Ruzov, A., Dunican, D. S., Prokhortchouk, A., Pennings, S., Stancheva, I., Prokhortchouk, E. and Meehan, R. R. (2004). Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. *Development* **131**, 6185-6194.

Ruzov, A., Savitskaya, E., Hackett, J. A., Reddington, J. P., Prokhortchouk, A., Madej, M. J., Chekanov, N., Li, M., Dunican, D. S., Prokhortchouk, E., Pennings, S. and Meehan, R. R. (2009). The non-methylated DNA-binding function of Kaiso is not required in early Xenopus laevis development. *Development* **136**, 729-738.

Valenta, T., Lukas, J., Doubravska, L., Fafilek, B. and Korinek, V. (2006). HIC1 attenuates Wnt signaling by recruitment of TCF-4 and beta-catenin to the nuclear bodies. *EMBO J.* **25**, 2326-2337.

Veeman, M. T., Slusarski, D. C., Kaykas, A., Louie, S. H. and Moon, R. T. (2003). Zebrafish prickle, a modulator of noncanonical Wnt/Fz signaling, regulates gastrulation movements. *Curr. Biol.* **13**, 680-685.

Weber, A., Marquardt, J., Elzi, D., Forster, N., Starke, S., Glaum, A., Yamada, D., Defossez, P. A., Delrow, J., Eisenman, R. N., Christiansen, H. and Eilers, M. (2008). Zbtb4 represses transcription of P21CIP1 and controls the cellular response to p53 activation. *EMBO J.* **27**, 1563-1574.

Development 136, 729-738 (2009) doi:10.1242/dev.025569

The non-methylated DNA-binding function of Kaiso is not required in early *Xenopus laevis* development

Alexey Ruzov[1,*], Ekaterina Savitskaya[2,*], Jamie A. Hackett[1], James P. Reddington[1], Anna Prokhortchouk[2], Monika J. Madej[1], Nikolai Chekanov[2], Minghui Li[3], Donncha S. Dunican[1], Egor Prokhortchouk[2], Sari Pennings[4] and Richard R. Meehan[1,†]

Mammalian forms of the transcription repressor, Kaiso, can reportedly bind methylated DNA and non-methylated CTGCNA motifs. Here we compare the DNA-binding properties of Kaiso from frog, fish and chicken and demonstrate that only the methyl-CpG-binding function of Kaiso is evolutionarily conserved. We present several independent experimental lines of evidence that the phenotypic abnormalities associated with xKaiso-depleted *Xenopus laevis* embryos are independent of the putative CTGCNA-dependent DNA-binding function of xKaiso. Our analysis suggests that xKaiso does not play a role in the regulation of either xWnt11 or Siamois, key signalling molecules in the Wnt pathway during *X. laevis* gastrulation. The major phenotypic defects associated with xKaiso depletion are premature transcription activation before the mid-blastula transition and concomitant activation of a p53-dependent cell-death pathway.

KEY WORDS: Evolution, Kaiso, MBT, Siamois, Methyl-CpG binding, *Xenopus laevis*

INTRODUCTION

Reduction in xKaiso protein levels by morpholino (xKMO) injection into *Xenopus laevis* embryos results in developmental delay during gastrulation relative to control morpholino (CMO)-injected embryos (Kim et al., 2004; Ruzov et al., 2004). xKMO morphants subsequently die during neurulation with all the hallmarks of apoptosis (Ruzov et al., 2004). Differing interpretations exist as to the molecular basis of the mutant phenotype. Our work suggests that xKaiso regulates general gene silencing before the mid-blastula transition (MBT) through its ability to bind methylated DNA via its zinc-finger domains (ZF1-3) (Ruzov et al., 2004). In this respect, it is notable that the ectopic gene expression profile in pre-MBT xKMO embryos corresponds to a subset of genes that are prematurely activated when levels of the maintenance methyltransferase, xDnmt1, are decreased (Ruzov et al., 2004). A different study (using the same xKaiso morpholino) was restricted to the analysis of potential gastrulation defects (stages 10-12), in which the same gastrula phenotype, corresponding to developmental delay and an open blastopore, was observed (Kim et al., 2004). Here it was suggested that xKaiso could also directly repress canonical and non-canonical Wnt gene targets (Siamois, Fos, Cyclin-D1, Myc and xWnt11) based on its ability to bind non-methylated CTGCNA sites that are present in target promoters (Kim et al., 2004; Park et al., 2005). These distinct reports suggested that xKaiso has bimodal gene regulatory roles during animal development; as a participant in an embryonic general transcription repression pathway and as a regulator of canonical and non-canonical Wnt-signalling pathways during gastrulation. These data also raise questions as to the underlying molecular pathology of the observed phenotypes. Do they result from Kaiso's role as a component of the xDnmt1/DNA methylation repression pathway in pre-MBT embryos and subsequent activation of apoptosis, or are the phenotypes due to its ability to specifically regulate the expression of genes such as Siamois and xWnt11 via defined non-methylated DNA binding sequences? One way to discriminate between these potentially differing roles in *X. laevis* development is to try to rescue the mutant phenotype with a Kaiso variant that can only bind methylated DNA and not CTGCNA-binding sites.

The original DNA-binding site selection experiments with mouse Kaiso under low stringency conditions identified a non-methylated DNA-binding motif, Hmat, with a conserved 6 bp core sequence CTGCNA that was first identified in the promoter of the human matrilysin gene (Daniel et al., 2002). Previously we had noted that xKaiso was not as robust as its mammalian counterparts in binding Hmat (Ruzov et al., 2004). We therefore undertook a characterisation of the DNA-binding properties of Kaiso in three species (zebrafish, frog and chicken) to determine if their methylated and non-methylated DNA-binding functions are conserved. In this study we demonstrate that the ZF1-ZF2 region of all three Kaiso homologues is sufficient for binding methylated DNA but that the ability to bind Hmat is not conserved. Zebrafish Kaiso is unable to bind Hmat or CTGCNA sequences present in the Siamois and xWnt11 promoters. Despite its reduced DNA-binding repertoire (compared with frog and human Kaiso), co-injection of dKaiso mRNA rescues developmental defects associated with xKMO morphants. This observation suggests that the reported CTGCNA-binding function of xKaiso does not have a key role during early *Xenopus laevis* development. In agreement with this observation, we did not observe ectopic activation of Siamois or xWnt11 expression in xKMO morphants. A global analysis of Kaiso occupancy in chromatin derived from human 293 cells also did not find any evidence for enrichment of CTGCNA-containing sequences. We propose that the main role of Kaiso in early *X. laevis* development is more restricted than previously suggested and intimately linked with the maintenance of transcriptional silencing before the onset of zygotic transcription at the MBT (Park et al., 2005; Ruzov et al., 2004).

[1]Human Genetics Unit, MRC, Western General Hospital, Edinburgh EH4 2XU, UK. [2]Center 'Bioengineering', 60-let Oktyabrya 7-1, Moscow, 117312, Russian Federation. [3]Shanghai Biochip Company, No.151 Libing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai, 201203, China. [4]Queen's Medical Research Institute, University of Edinburgh, Edinburgh EH16 4TJ, UK.

*These authors contributed equally to this work
†Author for correspondence (e-mail: r.meehan@hgu.mrc.ac.uk)

Accepted 15 December 2008

MATERIALS AND METHODS

Plasmids, recombinant proteins and reporter assays

The coding regions of *Danio rerio* and *Gallus gallus* Kaiso were amplified from genomic DNAs and cloned into pGEM-T-easy vector (Promega). A Kozak sequence and stop codon were introduced to the dKaiso plasmid used for in vitro transcription in rescue experiments. xKaiso ZF123 (GST-*x*ZF, aa 470-609) was described previously (Ruzov et al., 2004). The same region was cloned into a modified pet25A-6xHis-tag expression vector (Allen et al., 2006). The xBTB/POZ region (aa 2-120) of xKaiso was cloned into pGEX-4T1. dKaisoZF123 (aa 371-550) and gKaisoZF123 (aa 419-611) were cloned into both pGEX-6P-1 and pet25A-6xHis-tag. Full-length xKaiso was cloned into pGEX-6P-2. xKaisoZF12 (aa 470-558) and xKaisoZF23 (aa 523-608) were cloned into pGEX-6P-1 and pGEX-6P-3, respectively. gKaisoZF12 (aa 419-558) and gKaisoZF23 (aa 524-611) were cloned into the pGEX-6P1 vector. GST-fusion proteins were expressed in Rosetta-gami cells (Novagen) and purified on GST Sepharose. 6xHis-tag fusions were expressed in BL21-Codon Plus (DE3)-RIPL (Stratagene) and affinity purified on Ni-NTA Superflow (Qiagen) according to the manufacturer's instructions. The mKaiso construct was made by cloning the full-length mKaiso cDNA into the *Eco*RI site of the LZRS vector. The HA-tagged dKaiso construct was made by cloning into the C-terminal polylinker of CMV2-FLAG together with an HA epitope. The T7-tagged dKaiso and xKaiso expression constructs were made in pCGT7 by cloning into *Xba*I-*Bam*HI sites (Cazalla et al., 2005). These were expressed in 293T cells and T7-tagged proteins affinity purified (Cazalla et al., 2005). To obtain VP16 fusions zinc-finger domains of xKaiso (aa 447-635) and the HMG domain of xTcf3 were cloned into the *Eco*RI and *Xba*I sites downstream of the VP16 activation domain in pVP16 (Clontech). The –304 to +198 bp region of the mouse Tex19 promoter was cloned into the *Mlu*I and *Bgl*II sites of pGL3-basic (Promega) using primers with restriction site overhangs. S01234 and S constructs were previously described (Brannon et al., 1997). The constructs were in vitro methylated according to standard techniques. Luciferase reporters were transfected (Lipofectamine 2000) into 293T cells and analysed according to Dunican et al. (Dunican et al., 2008).

Electrophoretic mobility shift assay (EMSA) experiments

Binding reactions were as described, using 5% PAGE in 0.5 × TBE to resolve DNA-protein complexes with S (non-methylated), Sm (methylated-Sm) probes (Prokhortchouk et al., 2001), the human matrilysin (Hmat) oligo (Daniel and Reynolds, 1999), wild-type KCS-*Siamois* oligo (Park et al., 2005), wild-type *Xenopus Wnt11* oligo (Kim et al., 2004) and TCFbs oligo derived from *Xenopus Siamois* promoter: F catcagaatcATCAAAGgacctccc, R gggaggtccTTTGATgattctgatg. Purified GST proteins, 6xHis-tag proteins, T7-tagged proteins, *Escherichia coli* extracts containing GST-fusion proteins or in vitro translated myc-xTcf3 were used in EMSAs. After gel scanning on an FLA2000, signal quantification was performed using AIDA software.

Embryos and microinjections

Xenopus embryos were obtained from in vitro fertilised eggs. They were grown, staged and microinjected according to standard procedures (Stancheva and Meehan, 2000). At the two-cell stage, the embryos were injected into the animal half with 10-40 ng/cell of the xKMO or control morpholino (Gene-Tools), and/or 200-750 pg of sense capped RNA (dKaiso or myc-xKaiso mRNA) synthesized in vitro (T3/T7 Cap-Scribe kit, Boehringer) (Ruzov et al., 2004). Zebrafish Wik embryos, obtained from in house breeding, were maintained at 28°C as described previously (Detrich et al., 1999). Between 5 and 15 ng/embryo of dKaiso morpholino (ATATCAGCTTCAGTTTCGACATGCC) was injected according to Nasevicius and Ekker (Nasevicius and Ekker, 2000). A second MO, TGCAGAGCGACCCGTACAAATCCAC, was also used and gave similar phenotypes. For the rescue experiment equal volumes of dKMO and dKMO plus xKaiso mRNA (1 ng/nl) were injected. The phenotypes of surviving embryos (48) were scored after 24 hours. In experiments on apoptosis inhibition embryos were placed in 0.1XMMR containing 20 μM caspase-3 inhibitor Z-DEVD-FMK (R&D Systems) immediately after microinjection. xp53 morpholino was described in Cordenonsi et al. (Cordenonsi et al., 2003). All in situ hybridisations were performed according to published procedures (Hauptmann and Gerster, 1994). xWnt11 full-length cDNA was cloned into pGEMT-easy and used as a probe. xID2 probe was provided by Richard Harland (Liu and Harland, 2003).

Semi-quantitative and real-time RT-PCR

Semi-quantitative RT-PCR was performed as reported before (Ruzov et al., 2004) using published primers for *xWnt11* (Kim et al., 2004), *Siamois* (Park et al., 2005) and H4 (Ruzov et al., 2004). Quantitative real-time RT-PCR of *xWnt11*, *Siamois*, *Caspase7* and *Caspase9* was evaluated as follows (Houston et al., 2002). Primer sequences for Caspase7 and Caspase9 are available upon request. Total RNA was extracted with TriReagent (Sigma). Samples were reverse transcribed using random primers (Promega) and Superscript II RT (Invitrogen). Products were detected using SYBR Green PCR Mastermix (Applied Biosystems) and a PTC-200 cycler with a Chromo-4 detection system (MJ Research). Data were normalised relative to both GAPDH and H4 RNA, with comparable results. Error is expressed as s.e.m.

Chromatin immunoprecipitation (ChIP) assay and bisulfite sequencing

The chromatin immunoprecipitation (ChIP) assay was performed in an A6 *Xenopus* cell line according to (Dunican et al., 2008) using HA-tagged xKaiso or T7 tagged dKaiso with published ChIP primers for the *Siamois* promoter region (Park et al., 2005) or the Oct91 distal promoter region (Dunican et al., 2008). Bisulfite sequencing was performed according to standard procedures (Dunican et al., 2008).

HEK cells were used for genome-wide ChIP. Anti-HA tag polyclonal antibody (Bethyl Laboratories), ZFH6 rabbit polyclonal mKaiso antibody (Prokhortchouk et al., 2001), IgG or anti-T7 tag antibody (Novagene) were used. The transfection levels were checked by western blot hybridisation. ChIP DNA was amplified using anWGA4 kit (Sigma) and sequenced using the Genome Sequencer FLX System (Roche).

Genome-wide ChIP/sequencing data analysis

The ChIP DNA sequences were analysed using Perl, Blast and GS FLX Mapper software.

Initially PCR primer sequences were excluded from the analysis. Sequences obtained in the same experiment that were 97% or more homologous to each other and had the same 5′ end (with not more than 5 bp difference) were regarded as the same sequence that had become amplified during whole-genome amplification (WGA) PCR. All the remaining sequences were mapped on to the human genome (version 36.1) with a homology threshold of at least 95% throughout the whole length of the sequence. Only the unique genomic sequences were selected for further analysis. For all selected sequences the central positions were used as reference points for mapping onto the genome. If less than three such central positions were mapped onto a 1 kb segment of the genome the corresponding sequences were excluded from further analysis. In the case where central sequences positions from four different experiments (using different antibodies) were scored onto the same 1 kb region of the genome, the corresponding sequences were also excluded from the analysis. After this filtration protocol, the numbers of sequences for mKaiso ChIP with ZFH6 antibody, mKaiso ChIP with preimmune serum, dKaiso-HA ChIP with HA antibody and mock transfection with HA antibody experiments were 27,000, 72,000, 55,000 and 28,000, respectively. The 1 kb genomic regions anchored by the central sequences after filtration were analysed for the presence of CpG-rich regions or CTGCNA sites. Raw genome-wide ChIP/sequencing data are available upon request to Egor Prokhortchouk (Prokhortchouk@biengi.ac.ru).

CpG island array design

To design the CpG island array, CpG islands were selected from the human genome sequence (version 36.1) according to the following parameters: CpG island length >250 bp, expected/observed ratio >0.6, percentage of CpGs >50. There were 46,957 CpG islands chosen in total. These CpG islands were used by NimbleGene for the synthesis of 36,7802 isotermic oligonucleotides (38-70 bp each).

Genome-wide methylation status analysis

The MBD domain of human MBD2 was cloned into *Eco*RI, *Sal*I sites of pGEX 4T-1 vector to make MBD2 GST fusion construct. To prepare MBD2B-GST sepharose the MBD2B-GST fusion was purified on glutathione sepharose 4b (Amersham Biosciences, Piscataway, NJ) without elution. Fifty microlitres of sepharose saturated with GST-tagged MBD2B were incubated with 200 μl binding buffer (25 mM Hepes KOH, pH 7.5, 300 mM KCl, 12.5 mM $MgCl_2$, 10% glycerol, 1 mM DTT). ChIP HEK293 genomic DNA, 500 ng fragmented with an average size of 200-300 bp and ligated with adaptors (5′-GCGGTGACCCGGGAGATC-TGAATTC-3′ and 5′-GAATTCAGATC-3′), was used for binding with MBD2-GST-sepharose using a procedure adapted from Rauch and Pfeifer (Rauch and Pfeifer, 2005). Fifty nanograms of DNA was used as an input control. Briefly, DNA was incubated with MBD2 resin for 2 hours, washed three times with washing buffer (25 mM Hepes KOH, pH 7.5, 600 mM KCl, 12.5 mM $MgCl_2$, 10% glycerol, 1 mM DTT), eluted with elution buffer (25 mM Hepes KOH, pH 7.5, 1.5 M KCl, 12.5 mM $MgCl_2$, 10% glycerol, 1 mM DTT), purified using Qiaquick PCR purification kits (Qiagen, Valencia, CA), amplified and Cy3(Cy5) labelled. Equal amounts of MBD2B-GST bound and input labelled DNA were hybridised with our NimbleGene CpG island array for 40 hours according to the manufacturer's instructions. After hybridization, the array was processed with NimbleGene buffers and washes. The arrays were scanned on a GenePix scanner to measure Cy3 and Cy5 intensity. To identify signal peaks SignalMap software (NimbleGene) was used with a threshold 2. The CpG rich sequences obtained in the genome-wide ChIP/sequencing experiment were analysed for their correspondence (overlap) with the methylated CpG islands using a ChIP sequence/CpG island distance thresholds from 500 to 2000 bp.

RESULTS

The methyl-CpG-binding function of Kaiso is conserved

We identified full-length zebrafish (*Danio rerio*), *Fugu rubripes* and chicken (*Gallus gallus*) Kaiso homologues by database screening with human and *Xenopus* Kaiso sequences. In each case, the structural organisation was similar, with an N-terminal BTB/POZ domain and a C-terminal region containing three zinc fingers (ZF1-3). Alignment of these Kaiso protein sequences with their mammalian counterparts showed that BTB/POZ domains are very similar (53% identity, 82% similarity), as are ZF1 (85% identity, 90% similarity) and ZF2 (72% identity, 90% similarity). The ZF3 regions are much more variable, exhibiting 38% identity and 61% similarity overall (Fig. 1A). By contrast, the intervening region between the BTB and ZF domains exhibits low levels of overall similarity. Phylogenetic analysis reflected the evolutionary relationship between vertebrate species and suggested that fish Kaiso proteins form a group that is most distantly related from their tetrapod counterparts (see Fig. S1A,B in the supplementary material).

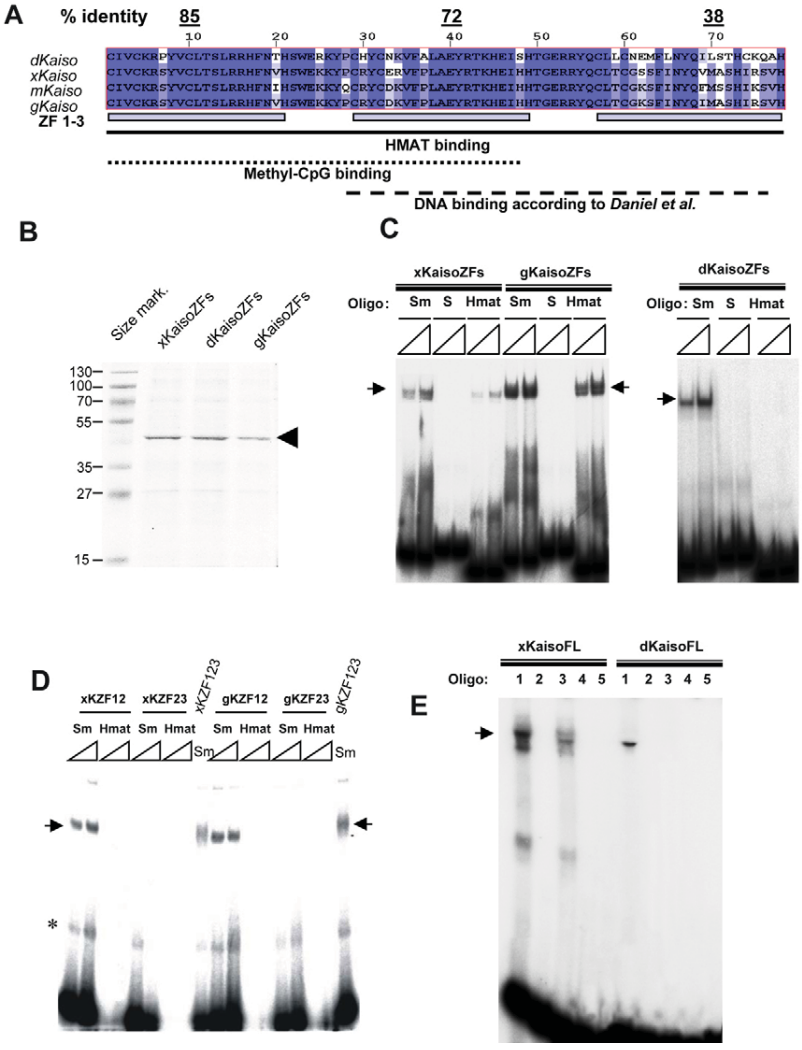


Fig. 1. The methyl-CpG binding activity of Kaiso is confined to ZF1 and 2. (**A**) Alignment of the zinc-finger domains of *Gallus*, *Xenopus*, mouse and *Danio* Kaiso proteins. The regions required for Hmat-dependent binding (solid line) and methyl-CpG-specific binding (dotted line) according to our results are compared with the previously published DNA-binding motif (dashed line) (Daniel et al., 2002). (**B**) SDS-PAGE showing the indicated purified 6xHis-tag ZF1-3 fusion proteins (arrow) used in the EMSAs. Size markers are on the left. (**C**) EMSA experiment with the indicated purified ZF1-3 proteins of xKaiso, dKaiso and gKaiso with methylated Sm, non-methylated S or human matrilysin (Hmat) probes in the presence of 2 μg pdIC competitor. Arrow indicates the Kaiso ZF-specific band shift. (**D**) EMSA using GST-ZF12 (xKZF12, gKZF12) and GST-ZF23 (xKZF23, gKZF23) deletion constructs from xKaiso and gKaiso, respectively, with Sm and Hmat oligos. xKaiso and gKaiso ZF domains (KZF123) were used with Sm probe as positive controls. (**E**) EMSA using eukaryotically expressed and affinity-purified full-length *Xenopus* (xKaisoFL) and *Danio* (dKaisoFL) Kaiso proteins with Sm (1), S (2), Hmat (3) and CTGCNA-containing probes from the promoter regions of Siamois (4) and xWnt11 (5).

The sequence variability in ZF3 is intriguing, as the ZF2-3 modules were reported to be necessary and sufficient for DNA binding (Daniel et al., 2002). We therefore determined the DNA-binding specificity of recombinant Kaiso-ZF1-3 modules in band shift assays. All the Kaiso proteins tested (*Xenopus*, zebrafish and chicken) were able to specifically bind the Sm oligo (3 MeCpGs) but not the S oligo, its non-methylated counterpart (Fig. 1C). By contrast, gKaiso had a similar affinity for the non-methylated Hmat sequence as for Sm, whereas xKaiso had a lower affinity for the Hmat oligo. Remarkably, dKaiso had no detectable binding to Hmat. This is probably due to sequence differences in ZF3. The same results were obtained using GST-ZF1-3 fusions (see Fig. S2A in the supplementary material). For a summary of the binding results see Fig. S2B in the supplementary material. It is notable that ZF3 in Kaiso proteins that can bind Hmat contains a core SHIR/KS sequence. This is replaced by THCKQ or THCKS in zebrafish and fugu, respectively (Fig. 1A), which may account for the lack of Hmat binding by dKaiso.

ZF1-2 is sufficient for binding to methylated DNA

Our comparative analysis demonstrates that the methyl-CpG-binding function is highly conserved in the three species tested, despite obvious sequence differences in ZF3. We investigated the ZF requirement for DNA binding using recombinant GST fusion proteins. We found that ZF1-2 from *Xenopus* and chicken is sufficient for binding to the Sm oligo (Fig. 1D). By contrast, an intact *Xenopus* and chicken ZF1-3 module is required to bind Hmat (Fig. 1C,D). This suggests that there is a correlation between the evolutionary conservation of ZF1-2 and the ability to bind methyl-CpGs. Full-length *Xenopus* and *Danio* Kaiso expressed and purified from 293 cells had the same DNA-binding specificity as their ZF1-3 counterparts (Fig. 1E); xKaiso could bind Sm (lane 1) but had a reduced affinity for Hmat (lane 3), whereas dKaiso could bind only Sm (lane 6). Our observations contrast with published work suggesting that ZF2-3 in mKaiso is sufficient for Hmat and Me-Sm binding (Daniel et al., 2002). The mKaiso analysis was based on EMSAs done under reduced stringency conditions with low amounts of competitor DNA and high amounts of fusion protein (Daniel et al., 2002). This may account for our differing conclusions; in our experiments, such EMSA conditions led to non-specific DNA binding (see Fig. 3E).

dKaiso is a methylation-specific repressor and is essential for zebrafish development

Apart from expression in *X. laevis* being limited to the animal pole, Kaiso mRNAs do not exhibit any regionally restricted expression pattern that would suggest Kaiso has a specific role in dorsoanterior specification (see Fig. S3A-C in the supplementary material). Given the unique DNA-binding properties of dKaiso, we tested whether it can repress transcription in a methyl-CpG dependent manner. We used a murine *Kaiso/Mecp2/Mbd2*$^{-/-}$ recipient cell line, which is defective in its ability to inhibit expression from methylated reporter plasmids (Filion et al., 2006). In these cells the methylated reporter plasmid was de-repressed to approximately 23% of a control non-methylated plasmid. Co-transfection of *Xenopus*, zebrafish or human Kaiso (*hKaiso*) expression plasmids resulted in enhanced repression of the methylated reporter, but not from the non-methylated control (Fig. 2A). Thus, dKaiso is a methyl-CpG repressor protein like its amphibian and mammalian counterparts. Since dKaiso cannot bind the non-methylated Hmat sequence, we determined whether it was essential for early zebrafish development using used a fluorescein-labelled morpholino (dKMO), which has the same specificity of action as its unlabelled counterpart. Embryos microinjected with dKMO at the 1- to 4-cell stage (Nasevicius and Ekker, 2000) were scored for survival and morphology after 24 and 48 hours of development (Fig. 2B,C). The dKMO morphants had significantly higher rates of embryo mortality at 48 hours: 89% compared with 17% for a non-inhibitory control morpholino, including any associated microinjection damage. Surviving dKMO morphants exhibited gross phenotypic defects, including microcephaly, that were coincident with the presence and dose of the morpholino. All the surviving control embryos (83%) went on to develop normally compared with only 2.5% of the dKMO morphants. Most of the remaining dKMO survivors at 48 hours

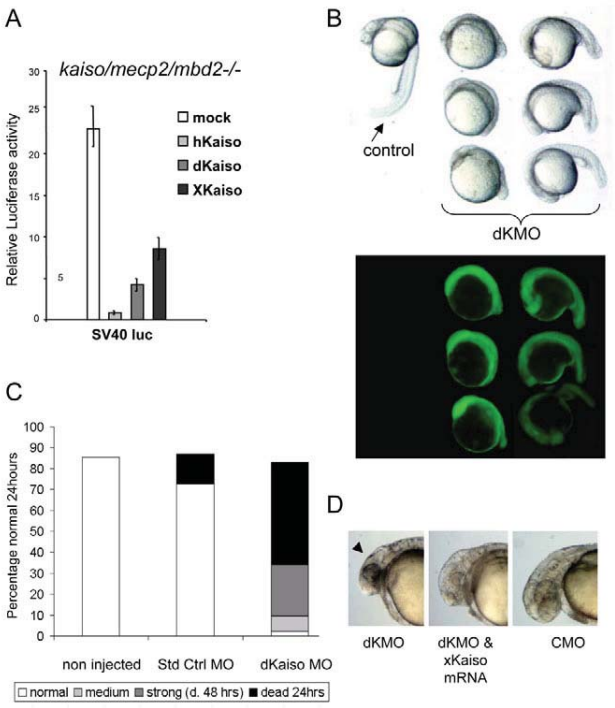


Fig. 2. *Danio rerio* Kaiso is a methyl-CpG-dependent repressor that is necessary for zebrafish development. (**A**) Methyl-CpG-dependent repression by dKaiso in a transient transfection assay. Kaiso expression constructs were co-transfected with a methylated SV40-luciferase reporter into mouse cells that are compromised in methyl-CpG-dependent transcriptional repression (*Kaiso/Mecp2/Mbd2*$^{-/-}$). The methylated SV40-luciferase reporter is repressed in the presence of dKaiso. The relative percentage (methylated reporter expression/nonmethylated reporter expression) is the average of at least three experiments. Human Kaiso (hKaiso) and xKaiso expression constructs were used as positive controls for methyl-CpG-dependent transcriptional repression. (**B**) The phenotypes of KMO-injected zebrafish embryos compared to control embryos 24 hours after fertilisation. Lower panel is an FITC image of the upper panel and shows that the severity of phenotypical defects correlates with the amount of injected fluorescein-labelled morpholino. (**C**) The percentages of normal embryos, embryos dead at 24 hours of development and embryos with strong (non-viable) and medium developmental abnormalities at 48 hours after fertilisation are shown for non-injected (n=259), standard control MO (Std Ctrl MO, n=62) and dKMO (n=182) embryos. (**D**) The microcephaly phenotype of dKMO morphants at 24 hours post-fertilisation (left panel) can be rescued by xKaiso mRNA (middle panel). A control (CMO-injected) embryo (right panel) at the same stage is shown for comparison.

(8.5%) exhibited developmental delay, axial defects and incomplete head formation. Although 2.5% were normal in appearance, these embryos displayed abnormal neural responses compared with controls (not shown) and represented a low-dose phenotype. These features are similar to the phenotypes associated with xKaiso or xDnmt1 depletion in *Xenopus* embryos and dDnmt1 depletion in zebrafish (Rai et al., 2006; Ruzov et al., 2004; Stancheva and Meehan, 2000). The range of dose-dependent phenotypes could be ameliorated by co-injection of xKaiso mRNA (Fig. 2D; see Fig. S4A,B in the supplementary material) and we observed similar phenotypes with a second non-overlapping dKaiso morpholino (not shown). We conclude that dKaiso is essential for zebrafish development.

Neither dKaiso or xKaiso can bind to CTGCNA sequences in the Siamois or xWnt11 promoters

Although it was unable to bind Hmat, it was possible that dKaiso could bind the CTGCNA-containing sequences associated with the Siamois and xWnt11 promoters (Park et al., 2005), and this feature underlies its essential function in zebrafish development. However, we could only detect binding by ZF1-3 (*Xenopus*, zebrafish and chicken) recombinant proteins to the Sm sequence and not the Siamois- and xWnt11-derived sequences (Fig. 3A,B; see Fig. S2C,D in the supplementary material), irrespective of using poly dI-dC or *E. coli* DNA as a non-specific competitor (data not shown). No binding to the promoter CTGCNA sequences was observed even with either a full-length GST-xKaiso fusion (see Fig. S2D in the supplementary material) or T7-tagged xKaiso or dKaiso (Fig. 1E). In competition experiments, unlabelled Sm oligo could efficiently compete with binding to itself, whereas Hmat was less efficient as a competitor of Sm binding to xZF1-3 (Fig. 3C,D). The competition experiments suggest that Hmat is 10-fold less efficient in binding xKaiso compared with Sm, whereas a 200-fold excess of either the Siamois- or xWnt11-derived sequences did not interfere with Sm binding (Fig. 3D). The specificity of the Kaiso/DNA complex is demonstrated by a supershift with an anti-His tag antibody (Fig. 3C). As a further test of stringency, we used a fixed amount of purified xZF1-3 protein in the presence of decreasing amounts of competitor DNA (2-10 ng) with labelled Hmat, Siamois, xWnt11 probes or a control duplex oligo with xTcf3- (non-CTGCNA) binding site. As expected, lowering the amount of competitor DNA led to increased formation of the xZF1-3/Hmat complex (Fig. 3E). However, we did not observe an interaction between Siamois- or xWnt11-derived sequences and xZF1-3 until the competitor was reduced to 10 ng; under these same conditions the unrelated xTcf3 target oligo also bound (Fig. 3E). On this basis we conclude that the Siamois- or xWnt11-derived sequences do not represent specific xKaiso-binding sites in vitro.

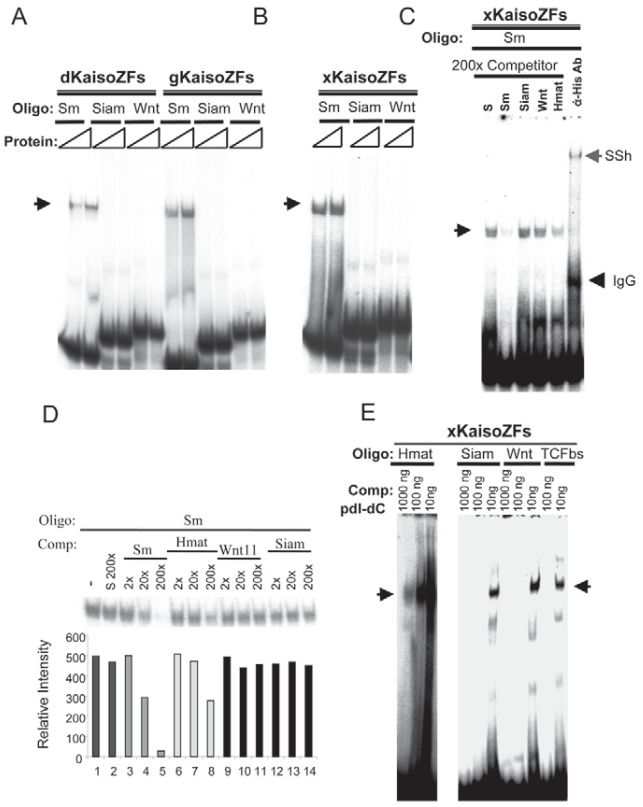


Fig. 3. xKaiso has no specific affinity for CTGCNA-binding sites in the *Siamois* and *xWnt 11* promoters. (**A**) EMSA experiment with the indicated purified 6xHis-ZF1-3 proteins dKaiso and gKaiso with methylated Sm probe and CTGCNA-containing probes derived from the promoter regions of Siamois (Siam) and xWnt11 (Wnt) in the presence of 2 μg pdIC competitor (**B**) Same experiment as in A but with xKaiso. (**C**) Competition experiment with 6xHis-ZF1-3 xKaiso protein under standard EMSA conditions with the Sm probe but with 200-fold excess of the following cold competitors: S, Sm, Siamois, xWnt11 and Hmat. Notice only the Sm and Hmat probes compete effectively. The last lane is a super-shift experiment in which an anti-His-tag antibody is included that shifts the 6xHis-Kaiso-specific complex. (**D**) Competition experiment as in C, with increasing amounts of cold competitors (2×, 20× and 200×). The signal quantification using AIDA software is shown below. Note that the Hmat oligo competes at least 10 times less efficiently than Sm. (**E**) EMSA with purified GST fusions of ZF1-3 domains of xKaiso with labelled Hmat, Siamois, xWnt11 and (non-CTGCNA) TCF3-binding site (TCFbs) probes. A fixed amount of protein and the indicated decreasing amount of pdIC competitor was used. Arrows indicate the shifted complex for each probe.

Kaiso preferentially associates with methylated CpGs rich sequences but not with CTGCNA sequences in vivo

To validate our in vitro EMSA results we performed a global analysis of Kaiso-binding sites in human HEK293 cells that were transiently transfected with either murine or *Danio* Kaiso expression plasmids. We sequenced and mapped chromatin-derived DNA fragments that were bound by these proteins and found a significant enrichment of CpG-rich sequences in both cases (Fig. 4A,C). A genome-wide analysis of DNA methylation in HEK293 cells showed that these CpG sequences bound by *Danio* and murine Kaiso are also enriched in methylated CpGs (see Fig. S5A in the supplementary material; data not shown). By contrast, our analysis did not detect an enrichment of CTGCNA sites in the chromatin fraction associated with mKaiso and dKaiso (Fig. 4B,D). These data do not exclude the possibility that mKaiso binds to a small number of CTGCNA-containing sequences in vivo but it does suggest that CTGCNA does not represent a general consensus sequence for mKaiso binding in vivo. The results are in remarkable agreement with our EMSA analysis and a ChIP experiment in which we could not detect any binding of transiently transfected dKaiso or xKaiso at the Siamois promoter in *Xenopus* A6 cells (Fig. 4F). By contrast, we could detect promoter occupancy by both dKaiso and xKaiso at the methylated distal promoter region of the Oct91 gene in the A6 cell line (Fig. 4E,F). An additional proof of the inability of xKaiso to

bind the Siamois promoter in vivo comes from luciferase reporter assays in which the xKaiso zinc fingers are fused to the VP16 activator domain (xKaisoZFVP16). xKaisoZFVP16 cannot activate transcription from a Siamois reporter (Fig. 4G, upper) but can activate transcription from a methylated Tex19 reporter 2.7 times (Fig. 4G, lower). It is important to note that a VP16 fusion with the xTcf3 DNA-binding HMG domain activated the Siamois-driven luciferase reporter 5-fold in the same set of experiments (Fig. 4G). These experiments support the view that xKaiso has a preference towards methylated DNA and not for the CTGCNA sequence present in the Siamois promoter.

dKaiso can rescue Kaiso-depleted *Xenopus laevis* embryos

The unique DNA-binding and transcriptional repressor specificity of dKaiso provides us with a perfect tool to test if the ability to bind methylated DNA is sufficient to rescue xKMO morphants. Although we previously found that 10 ng morpholino is sufficient to elicit a phenotype, we injected a high dose (40 ng), as used by Kim and colleagues, into *Xenopus* two-cell-stage embryos (Kim et al., 2004; Ruzov et al., 2004). In xKMO morphants the dorsal lip was less prominent at stage 10 and by stage 12 there was an extended open blastopore (see also Fig. 7A). This resulted in gastrulation delay followed by the characteristic appearance of white apoptopic cells near the edge of the open blastopore by stage 15 (Fig. 5A, xKMO and Fig. 7A). This gastrulation phenotype is identical to that reported by Kim and colleagues (however, they did not report on later stage phenotypes) with the same morpholino sequence and dose (Kim et al., 2004) and can be rescued by co-injection of either xKaiso, dKaiso or hKaiso mRNAs (Fig. 5B) (A.R. and R.R.M., unpublished). We found at stage 14 (neurulation) that a high rate of mortality (80%) occurred in the xKMO morphants compared with controls (Fig. 5B), and the embryos were developmentally arrested with, as noted before, a phenotype resembling that of xDnmt1-depleted embryos (Ruzov et al., 2004; Stancheva and Meehan, 2000). Embryos injected at the two-cell stage with dKaiso RNA and xKMO increased the proportion of normally gastrulating embryos (stage 12), from just over 10% to more than 45%. By stage 14 (neurulation) the mortality rate was reduced to 18% compared with more than 80% for xKMO-only injected embryos (Fig. 5A,B, xKMO+dKaiso). The presence of dKaiso mRNA enabled nearly half of the morphants (46%) to successfully complete gastrulation/neurulation compared with none for the xKMO alone (Fig. 5B). By tadpole stage, the rate of survival for dKaiso/xKMO-injected embryos was reduced, but it is notable that a few phenotypically normal embryos could develop even though the morpholino was still present (Fig. 5A,B). These rescued embryos differed in appearance from the few arrested xKMO morphants that survived to late stages (Fig. 5A). Importantly, the rescuing capacity of dKaiso mRNA was similar to that observed with wild-type (but not a DNA-binding mutant) hKaiso mRNA, which binds the Sm and Hmat sequences (Fig. 5B) (Ruzov et al., 2004). These experiments strongly suggest that the capacity to bind CTGCNA-containing sequences is not obligatory for rescuing xKMO-injected embryos; instead, the ability to bind methylated DNA correlates with the capacity to rescue early developmental deficits. This observation is in line with the proposed role of Kaiso as a transcriptional silencer in pre-MBT embryos (Ruzov et al., 2004). The ability of dKaiso to rescue the xKMO morphants is incompatible with a model in which a CTGCNA-binding activity is necessary to target the repression function of Kaiso to Wnt-signalling target genes (Kim et al., 2004; Park et al., 2005).

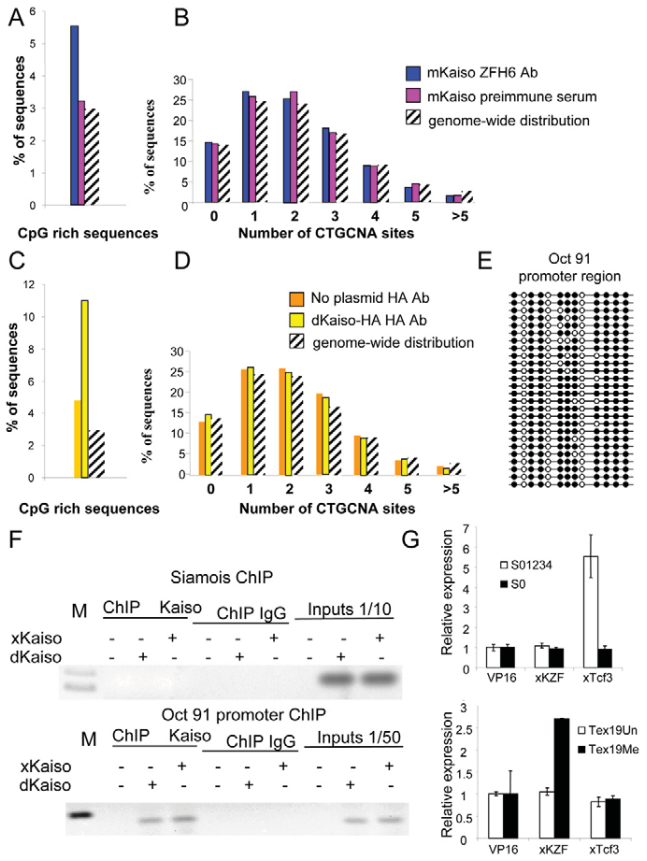


Fig. 4. Kaiso preferentially interacts with methylated CpGs but not with CTGCNA sequences in vivo. (**A-D**) The results of genome-wide ChIP/sequencing experiments in HEK293 transiently transfected with mouse Kaiso (mKaiso) and HA-tagged *Danio* Kaiso (dKaiso-HA). The ChIPs were performed using anti-mKaisoZF or anti-HA-tag antibodies with additional controls using preimmune serum (control for mKaiso ChIP) or anti-HA-tag antibody on non-transfected cells (control for dKaiso-HA experiment). The DNA obtained in the ChIP was amplified and 454 sequenced. After initial data filtering all the remaining sequences were mapped on to the human genome, susbsequently 1 kb regions in the vicinity of the ChIP sequences were analysed for the presence of either CpG-rich regions or CTGCNA sites. The ChIP sequences for both the mKaiso and dKaiso-HA experiments were enriched in CpG-rich regions in comparison to either to preimmune serum or anti-HA-tag antibody on non-transfected cells, respectively (A,C), but not in CTGCNA sites (B,D). The data were normalised with respect to the genome-wide distribution of CpG rich regions and CTGCNA sites as shown. (**E**) Diagram indicating the DNA methylation status of the Oct 91 distal promoter fragment used for ChIP in (F) in A6 cells. Filled circles represent methylated, and empty circles non-methylated, CpGs. (**F**) ChIP experiment using transiently transfected *Xenopus* HA-tagged Kaiso (xKaiso) and T7 tagged *Danio* Kaiso (dKaiso) on A6 cells. Both xKaiso and dKaiso bind to the heavily methylated distal region of the Oct91 gene, but do not show any detectable binding to the Siamois promoter under the same experimental conditions. IgG was used as an antibody control. 1/10 and 1/50 of inputs are loaded for the Siamois and Oct91 experiments, respectively. (**G**) The xKaiso ZF domain VP16 fusion (xKZF) does not activate transcription of a Siamois-driven luciferase reporter (S01234) but does activate transcription from a methylated Tex19 promoter (Tex19Me). The xTcf3 HMG domain fusion (xTcf3) activates transcription from the Siamois reporter 5.5 times. A Siamois luciferase reporter containing mutated xTcf3-binding sites (S0) and an unmethylated Tex19 promoter reporter (Tex19Un) were used as controls.

Inhibition of apoptosis in KMO embryos results in their successful gastrulation

xKMO morphants exhibit a developmental delay in closing the blastopore relative to control morpholino injected or wild-type embryos and subsequently die during neurulation with all the hallmarks of apoptosis (Fig. 7A) (Ruzov et al., 2004). We reasoned that inhibitors of the apoptotic pathway would reduce the high rates of lethality associated with the high-dose xKMO morphants. This would allow us to determine whether the resulting embryos exhibit a phenotype that would be indicative of ectopic Wnt signalling function. Incubation of wild-type embryos with the caspase-3 inhibitor, Z-DEVD-FMK, did not interfere with normal development (Fig. 6A). In high-dose xKMO morphants, the presence of Z-DEVD-FMK inhibited apoptosis and allowed up to one-third of embryos to progress through to the equivalent of tadpole stage (St. 38) (Fig. 6A; see Fig. S6 in the supplementary material). However, even though these embryos still underwent developmental delay, they gastrulated successfully but exhibited a short axis phenotype at later stages with obvious eye defects (Fig. 6A). We also suppressed the apoptotic effect of xKaiso depletion by directly inhibiting xp53 function by co-injecting a well-characterised xp53 morpholino (xp53MO) (Cordenonsi et al., 2003). The double xKMO/xp53MO morphants also underwent developmental delay but eventually underwent blastopore closure (Fig. 6B,C). The survival rate of these double morphants was high (90%) but none developed normally (Fig. 6C; see Fig. S6 in the supplementary material). These morphants did not exhibit an axis duplication phenotype that might be indicative of hyperactivation of downstream Wnt-signalling target genes (Tao et al., 2005). More importantly, these experiments imply that much of the phenotypic defects associated with xKaiso depletion are associated with a general failure to complete proper gastrulation/neurulation along with a concomitant activation of a p53-mediated cell-death pathway.

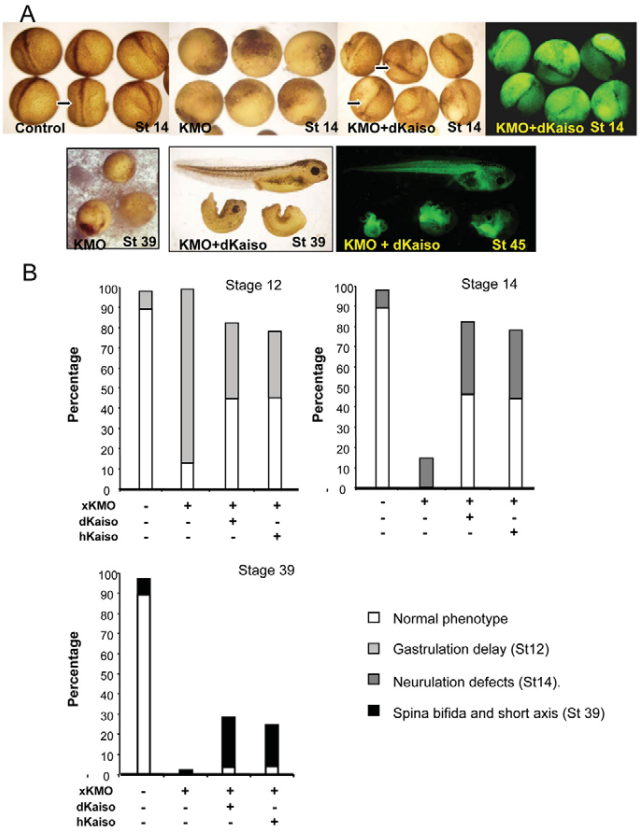


Fig. 5. dKaiso can rescue Kaiso-depleted *Xenopus laevis* embryos to the same extent as its human counterpart. (**A**) The phenotypes of uninjected control (*n*=150), KMO (*n*=53) and xKMO co-injected with dKaiso RNA (*n*=84) embryos (KMO+dKaiso). Development stages are indicated. FITC image of two pictures are presented as well as an injection control; arrow indicates neural fold. Notice that even at the later stage (St. 39), when there are reduced numbers of survivors, the xKMO morphants are arrested whereas the rescued embryos can form complete tadpoles or attenuated tadpoles that differ in appearance from the xKMO morphant. (**B**) Bar graphs showing the percentages of normal embryos and embryos with developmental defects in the rescue experiments using xKMO co-injected with dKaiso or human Kaiso (hKaiso) RNA. Dead embryos are not included. The stages of development are indicated.

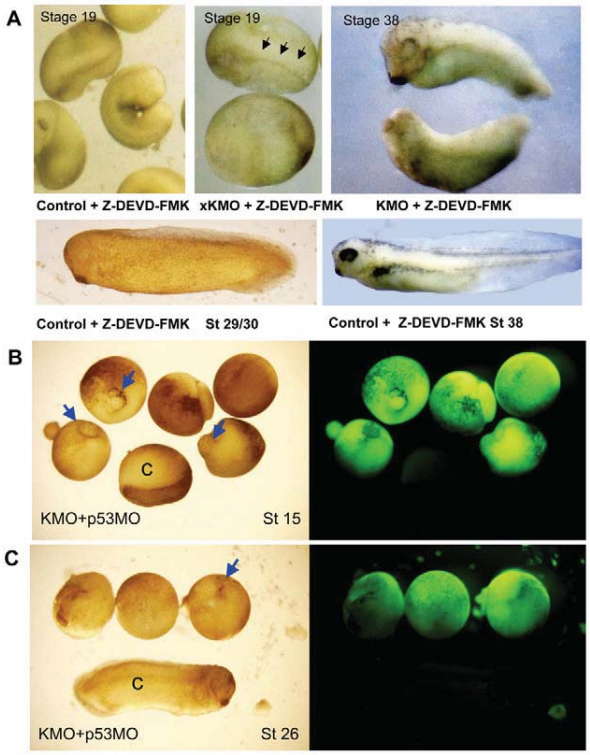


Fig. 6. Inhibition of apoptosis in KMO embryos results in their successful gastrulation. (**A**) The presence of a caspase inhibitor, Z-DEVD-FMK, prevents apoptosis in xKMO morphants and allows development to proceed. The rescued embryos can complete gastrulation but neurulation is impaired and they exhibit development delay compared to control embryos. Arrows indicate a poorly developed neural fold in the rescued xKMO morphant at stage 19. The rescued xKMO morphants do not show evidence of axis duplication (indicating no hyper β-catenin activation during gastrulation) at stage 38. In addition they exhibit developmental delay and axis defects that result from poor neurulation. Control embryos incubated with Z-DEVD-FMK are phenotypically normal. (**B**,**C**) Phenotypes of embryos co-injected with xKMO together with an xp53 morpholino (p53MO) are presented at stages 15 and 26. Uninjected control embryos are also shown (C in figures). Note the completion of delayed gastrulation in KMO/p53MO embryos (blue arrow points to the blastopores).

xWnt11 and Siamois are not mis-expressed in xKMO morphants

As result of our phenotypic analysis and the inability of xZF1-3 to bind the Siamois and xWnt11 probes, we decided to verify if these genes were aberrantly expressed in xKMO morphants, as reported previously (Kim et al., 2004; Park et al., 2005). Although xKMO embryos are delayed in gastrulation compared to wild-type embryos (Fig. 7A), we did not observe upregulation of Siamois and xWnt11 levels in xKMO morphants by semi-quantitative RT-PCR (Fig. 7B). Real-time quantitative RT-PCR suggests that Siamois levels are actually reduced by 30% in xKMO stage 10 morphants, which may reflect the developmental lag (Fig. 7A,C). By contrast, there were no significant differences in xWnt11 levels at the equivalent of stage 12 (Fig. 7C). In the same sets of embryos Caspase-7 and Caspase-9 RNA levels were increased 2.5-3.5 times in stage 10 and 12 xKMO embryos, respectively, compared with controls, which is in line with activation of a programmed cell-death pathway (Fig. 7C). We also used whole-mount RNA in situ analysis to compare the expression of xWnt11 in wild-type and xKMO pre-MBT embryos. Unlike previous reports (Kim et al., 2004), we could not find differences in expression: the maternal xWnt11 transcript was present at similar levels in both types of embryo (Fig. 7D; see Fig. S7A in the supplementary material). A control in situ with xID2, identified in a screen for genes that are mis-expressed in KMO embryos (Ruzov et al., 2004), confirms that it is upregulated in xKMO morphants (Fig. 7D; see Fig. S7B in the supplementary material). XWnt11 expression can be induced by the mesoderm-specific transcriptional activator Xbra (Tada and Smith, 2000), but we have not observed ectopic induction of Xbra in xKMO embryos, suggesting that xKaiso depletion does not result in the direct or indirect activation of xWnt11 (D.S.D., unpublished). We conclude that it is unlikely that xKaiso strongly influences the expression of xWnt11 and Siamois either before the MBT or during subsequent gastrulation.

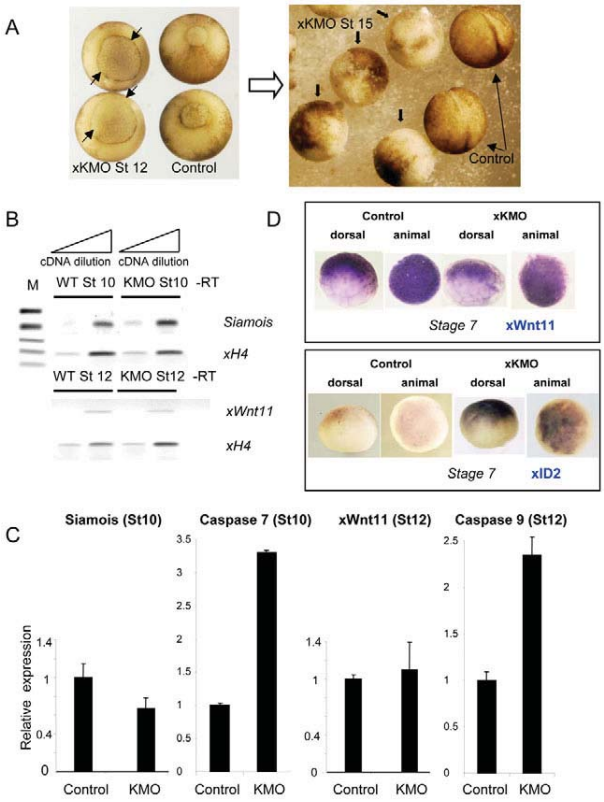


Fig. 7. The Wnt signalling pathway is not activated in xKMO morphants. (**A**) The xKMO morphants exhibit a delay in closing the blastopore (short arrows) compared with wild-type embryos. This phenotype appears in 85-90% of the embryos and is identical to that presented by McCrea and colleagues (Kim et al., 2004; Park et al., 2005). By stage 15 (neurulation), the xKMO morphants (downward arrow) cannot form a neural fold; they are apoptotic and are shedding cells through the open blastopore. Control embryos are shown with a proper neural fold (long arrows). (**B**,**C**) Neither *Siamois* nor *xWnt11* are ectopically activated in xKMO morphants when assayed by semi-quantitative RT-PCR according to Park et al. (Park et al., 2005) or real-time PCR relative to a histone H4 control at stage 10 (Siamois) or 12 (xWnt11). Caspase7 and Caspase9 expression is activated in 2.5-3.5 times compared to control in the same sets of KMO embryos at stages 10 and 12, respectively. (**D**) Whole-mount RNA in situ analysis demonstrates that xWnt11 is not prematurely activated in pre-MBT xKMO morphants in comparison to a control transcript, xID2, that is activated prematurely (Ruzov et al., 2004).

DISCUSSION

We find that the methyl-CpG-binding function of the Kaiso transcription repressor is highly conserved in three species; by contrast, the non-methylated Hmat-binding affinity varied from zero activity and weak binding for dKaiso and xKaiso, respectively, to levels comparable with methyl-CpG binding for gKaiso. In addition, our localisation of the Kaiso methyl-CpG-binding domain to ZF1-2 conflicts with previously published work that defined ZF2-3 as sufficient for both methyl-CpG-dependent and CTGCNA-based binding (Daniel et al., 2002). This previous observation may be attributable to differences in the stringency of the DNA-binding assays employed. However, the evolutionary conservation of the first two Kaiso ZF domains is in step with the experimental outcome of our DNA-binding analysis, which indicated that the methyl-CpG binding function resides in ZF1-2. Two other Kaiso-like proteins (ZBTB4 and ZBTB38) in mice also have a strong methyl-CpG-binding function, but only ZBTB4 has a weak affinity for the Hmat sequence, whereas ZBTB38 does not (Filion et al., 2006). Based on this, we come to a broad conclusion that the primary DNA-binding activity of all members of the Kaiso-like family (Kaiso, ZBTB4 and ZBTB38) is for methyl-CpGs, which can act as ligands in chromatin for the transcription repression function associated with these proteins (Filion et al., 2006; Prokhortchouk et al., 2001; Ruzov et al., 2004). Our global analysis of Kaiso-binding sites in HEK293 cells also supports the view that they are associated with methylated CpGs and not CTGCNA sites. In addition it has recently been shown that Kaiso represses methylated tumour suppressor genes and can bind in a methylation-dependent manner to the CDKN2A gene in human colon cancer cell lines (Lopes et al., 2008).

The DNA-binding ability of Kaiso is primarily dependent on its three zinc fingers (ZF1-3), which belong to the C2H2 class. It has been proposed that each ZF can be regarded as an independent DNA-binding module, and an additional ZF in an array specifies three base pairs of adjacent, but discrete, subsites (Choo and Klug, 1997). This would predict a 9 bp binding site for Kaiso if each ZF bound equally to its respective subsite. An inference of the requirement for all three ZFs to bind Hmat is that its true recognition sequence may correspond to a 9 bp sequence and not 6 bp as previously reported (Daniel et al., 2002). The requirement for ZF1-2 for binding to the Sm oligo is consistent with a minimum recognition sequence of 6 bp that contains two methyl-CpGs (Prokhortchouk et al., 2001). Comparison of the three CTGCNA substrates used (Hmat, Siamois and xWnt11) shows that they are

flanked by distinct sequences that probably account for their differential binding to gKaiso and xKaiso. Notably, like xKaiso, gKaiso cannot bind the CTGCNA-containing sequences in the xWnt11 and Siamois promoter regions (Fig. 3) despite its high affinity for Hmat. This suggests that the non-methylated binding ability of gKaiso is restricted to one or a few unique sequences that have yet to be identified. Our analysis suggests that CTGCNA sequences in the promoters of Siamois and xWnt11 are not high-affinity Kaiso-binding sites, which would be incompatible with the proposed role of xKaiso as a direct repressor of these genes (Park et al., 2005).

We also show, that like its amphibian counterpart, dKaiso is essential for early zebrafish development. The phenotypes of dKMO-injected embryos, the percentages of abnormal embryos and the developmental stages in which these embryonic defects are detected resemble those of xKMO-injected *Xenopus* embryos (Ruzov et al., 2004). In our previous study we showed that xKaiso is required for genome-wide transcription silencing in embryos before the MBT. In zebrafish the MBT begins at cycle 10 and, as in amphibians, is characterised by cell-cycle lengthening, loss of cell synchrony, appearance of cell motility and activation of transcription (Kane and Kimmel, 1993). Noting the similarity of the *Danio rerio* and *Xenopus* Kaiso loss-of-function phenotypes, we can hypothesise that in both cases mis-regulation of similar mechanisms leads to close phenotypic abnormalities. Another conclusion from the comparison of zebrafish and *Xenopus* KMO embryos and from the dKaiso ability to rescue the xKMO phenotype, is that the main phenotypical features of xKMO-injected embryos are not dependent on CTGCNA-binding ability, which is not conserved between zebrafish and *Xenopus*. The observation that mice do not have an extended period of transcriptional silencing during early development may partly explain the absence of an embryonic lethal phenotype in Kaiso null mice (Prokhortchouk et al., 2006). These mutant mice also do not exhibit mis-regulation of candidate target genes such as Wnt11.

In parallel with the mouse study, we did not observe the reported mis-regulation of the Wnt signalling pathway in xKMO stage 8 or stage 10 morphants. Indeed, inhibition of the cell death pathway in xKMO morphants by incubation with apoptotic inhibitors does not uncover any gross axis duplication phenotypes, which is also indicative that β-catenin signalling is not ectopically activated under these conditions. A potential role for xKaiso in Wnt signalling was initially suggested by non-physiological experiments in which xKaiso and Wnt-signalling components are overexpressed. This is probably mediated by Kaiso/TCF3 interactions and not via CTGCNA-binding sites (Kim et al., 2004; Park et al., 2005; Ruzov et al., 2009). In a companion paper, we demonstrate that xKaiso interacts directly with TCF3/4 and thereby masks their HMG DNA-binding domains (Ruzov et al., 2009). As a result, overexpression of xKaiso inhibits the ability of β-catenin to mediate transcription activation through xTcf3 by displacing it from a target promoter. This mode of action can account for the inhibition of dorsal axis formation when Kaiso mRNA is co-injected with β-catenin mRNA into the ventral marginal region of four-cell-stage embryos (Park et al., 2005).

We thank Josh Brickman, Patricia Yeyati and Nick Hastie for helpful comments and corrections during manuscript preparation and Nick Hastie for general support. Work in the R.R.M. lab (R.R.M., D.S.D., A.R. and J.A.H.) is supported by the MRC. We thank Richard Harland, Vladimir Korinek, Pierre McCrea, Randall Moon and Ralph Rupp for reagents, Mike McGrew and Helen Sang for chick embryos, Nicola Grey and Bill Richardson for *Xenopus* embryos and members of the Chromosome and Gene expression group for technical advice. E.P. and A.P. were supported by Russian Academy of Sciences program grant "Molecular and Cellular Biology", RFBR 06-04-49216-a and CRDF RB-2808. M.L. was supported by CR10-A-04 and 08410708700. Deposited in PMC for release after 6 months.

Supplementary material

Supplementary material for this article is available at http://dev.biologists.org/cgi/content/full/136/5/729/DC1

References

Allen, M. D., Grummitt, C. G., Hilcenko, C., Min, S. Y., Tonkin, L. M., Johnson, C. M., Freund, S. M., Bycroft, M. and Warren, A. J. (2006). Solution structure of the nonmethyl-CpG-binding CXXC domain of the leukaemia-associated MLL histone methyltransferase. *EMBO J.* **25**, 4503-4512.

Brannon, M., Gomperts, M., Sumoy, L., Moon, R. T. and Kimelman, D. (1997). A beta-catenin/XTcf-3 complex binds to the siamois promoter to regulate dorsal axis specification in Xenopus. *Genes Dev.* **11**, 2359-2370.

Cazalla, D., Sanford, J. R. and Caceres, J. F. (2005). A rapid and efficient protocol to purify biologically active recombinant proteins from mammalian cells. *Protein Expr. Purif.* **42**, 54-58.

Choo, Y. and Klug, A. (1997). Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.* **7**, 117-125.

Cordenonsi, M., Dupont, S., Maretto, S., Insinga, A., Imbriano, C. and Piccolo, S. (2003). Links between tumor suppressors: p53 is required for TGF-beta gene responses by cooperating with Smads. *Cell* **113**, 301-314.

Daniel, J. M. and Reynolds, A. B. (1999). The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. *Mol. Cell. Biol.* **19**, 3614-3623.

Daniel, J. M., Spring, C. M., Crawford, H. C., Reynolds, A. B. and Baig, A. (2002). The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res.* **30**, 2911-2919.

Detrich, H. W., 3rd, Westerfield, M. and Zon, L. I. (1999). Overview of the Zebrafish system. *Methods Cell Biol.* **59**, 3-10.

Dunican, D. S., Ruzov, A., Hackett, J. A. and Meehan, R. R. (2008). xDnmt1 regulates transcriptional silencing in pre-MBT Xenopus embryos independently of its catalytic function. *Development* **135**, 1295-1302.

Filion, G. J., Zhenilo, S., Salozhin, S., Yamada, D., Prokhortchouk, E. and Defossez, P. A. (2006). A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol. Cell. Biol.* **26**, 169-181.

Hauptmann, G. and Gerster, T. (1994). Two-color whole-mount in situ hybridization to vertebrate and Drosophila embryos. *Trends Genet.* **10**, 266.

Houston, D. W., Kofron, M., Resnik, E., Langland, R., Destree, O., Wylie, C. and Heasman, J. (2002). Repression of organizer genes in dorsal and ventral Xenopus cells mediated by maternal XTcf3. *Development* **129**, 4015-4025.

Kane, D. A. and Kimmel, C. B. (1993). The zebrafish midblastula transition. *Development* **119**, 447-456.

Kim, S. W., Park, J. I., Spring, C. M., Sater, A. K., Ji, H., Otchere, A. A., Daniel, J. M. and McCrea, P. D. (2004). Non-canonical Wnt signals are modulated by the Kaiso transcriptional repressor and p120-catenin. *Nat. Cell Biol.* **6**, 1212-1220.

Liu, K. J. and Harland, R. M. (2003). Cloning and characterization of Xenopus Id4 reveals differing roles for Id genes. *Dev. Biol.* **264**, 339-351.

Lopes, E. C., Valls, E., Figueroa, M. E., Mazur, A., Meng, F. G., Chiosis, G., Laird, P. W., Schreiber-Agus, N., Greally, J. M., Prokhortchouk, E. et al. (2008). Kaiso contributes to DNA methylation-dependent silencing of tumor suppressor genes in colon cancer cell lines. *Cancer Res.* **68**, 7258-7263.

Nasevicius, A. and Ekker, S. C. (2000). Effective targeted gene 'knockdown' in zebrafish. *Nat. Genet.* **26**, 216-220.

Park, J. I., Kim, S. W., Lyons, J. P., Ji, H., Nguyen, T. T., Cho, K., Barton, M. C., Deroo, T., Vleminckx, K., Moon, R. T. et al. (2005). Kaiso/p120-catenin and TCF/beta-catenin complexes coordinately regulate canonical Wnt gene targets. *Dev. Cell* **8**, 843-854.

Prokhortchouk, A., Hendrich, B., Jorgensen, H., Ruzov, A., Wilm, M., Georgiev, G., Bird, A. and Prokhortchouk, E. (2001). The p120 catenin partner Kaiso is a DNA methylation-dependent transcriptional repressor. *Genes Dev.* **15**, 1613-1618.

Prokhortchouk, A., Sansom, O., Selfridge, J., Caballero, I. M., Salozhin, S., Aithozhina, D., Cerchietti, L., Meng, F. G., Augenlicht, L. H., Mariadason, J. M. et al. (2006). Kaiso-deficient mice show resistance to intestinal cancer. *Mol. Cell. Biol.* **26**, 199-208.

Rai, K., Nadauld, L. D., Chidester, S., Manos, E. J., James, S. R., Karpf, A. R., Cairns, B. R. and Jones, D. A. (2006). Zebra fish Dnmt1 and Suv39h1 regulate organ-specific terminal differentiation during development. *Mol. Cell. Biol.* **26**, 7077-7085.

Rauch, T. and Pfeifer, G. P. (2005). Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer. *Lab. Invest.* **9**, 1172-1180.

Ruzov, A., Dunican, D. S., Prokhortchouk, A., Pennings, S., Stancheva, I., Prokhortchouk, E. and Meehan, R. R. (2004). Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. *Development* **131**, 6185-6194.

Ruzov, A., Hackett, J. A., Prokhortchouk, A., Reddington, J. P., Madej, M. J., Dunican, D. S., Prokhortchouk, E., Pennings, S. and Meehan, R. R. (2009). The interaction of xKaiso with xTcf3: a revised model for integration of epigenetic and Wnt signalling pathways. *Development* **136**, 723-727.

Stancheva, I. and Meehan, R. R. (2000). Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos. *Genes Dev.* **14**, 313-327.

Tada, M. and Smith, J. C. (2000). Xwnt11 is a target of Xenopus Brachyury: regulation of gastrulation movements via Dishevelled, but not through the canonical Wnt pathway. *Development* **127**, 2227-2238.

Tao, Q., Yokota, C., Puck, H., Kofron, M., Birsoy, B., Yan, D., Asashima, M., Wylie, C. C., Lin, X. and Heasman, J. (2005). Maternal wnt11 activates the canonical wnt signaling pathway required for axis formation in Xenopus embryos. *Cell* **120**, 857-871.

Development 135, 1295-1302 (2008) doi:10.1242/dev.016402

xDnmt1 regulates transcriptional silencing in pre-MBT *Xenopus* embryos independently of its catalytic function

Donncha S. Dunican[1,2], Alexey Ruzov[1,3], Jamie A. Hackett[1] and Richard R. Meehan[1,2,*]

We previously reported that the maintenance cytosine methyltransferase xDnmt1 is essential for gene silencing in early *Xenopus laevis* embryos. In the present study, we show that silencing is independent of its catalytic function and that xDnmt1 possesses an intrinsic transcription repression function. We show that reduction of xDnmt1p by morpholino (xDMO) injection prematurely activates gene expression without global changes in DNA methylation before the mid-blastula transition (MBT). Repression of xDnmt1p target genes can be reimposed in xDMO morphants with an mRNA encoding a catalytically inactive form of human DNMT1. Moreover, target gene promoter analysis indicates that silencing is not reliant on dynamic changes in DNA methylation. We demonstrate that xDnmt1 can suppress transcription activator function and can be specifically localised to non-methylated target promoters. These data imply that xDnmt1 has a major silencer role in early *Xenopus* development before the MBT as a direct transcription repressor protein.

KEY WORDS: DNA methylation, MBT, *Xenopus*

INTRODUCTION

DNMT1 is a multi-domain protein with an N-terminal regulatory domain and a C-terminal catalytic domain (Goll and Bestor, 2005). The enzymatic function of DNMT1 is necessary to maintain and perpetuate DNA methylation patterns at CpGs laid down by de novo methyltransferases in response to developmental cues. DNA methylation in mammals is essential for transcriptional silencing of transposons, regulation of many imprinted genes and the maintenance of X-inactivation in female somatic cells (Goll and Bestor, 2005). Methylated CpG pairs can repress transcription of adjacent genes either by directly interfering with nuclear factor site recognition or, indirectly by binding methyl-CpG specific binding proteins (Klose and Bird, 2006). Altered patterns of gene expression (including single copy genes, imprinted genes and transposons) occurs in hypomethylated (*dnmt1*) mutant mouse embryonic fibroblasts (Jackson-Grusby et al., 2001). Nonetheless, many tissue-specific gene promoters are hypo-methylated and are not expressed in early mouse or *Xenopus* embryos (Walsh and Bestor, 1999; Stancheva et al., 2002), implying that additional silencing mechanisms may be operative. A screen of *dnmt1*$^{n/n}$ fibroblast cells also noted that a high proportion of mis-expressed genes are transcribed from CpG-island promoters that are constitutively unmethylated (Lande-Diner et al., 2007).

A decrease in DNMT1 levels results in early embryonic lethality in mouse, frog and zebrafish, probably owing to multiple defects, including activation of a cell death pathway (Li et al., 1992; Stancheva and Meehan, 2000; Jackson-Grusby et al., 2001; Stancheva et al., 2001; Rai et al., 2006). Recently a mutant mouse with a catalytically inactive form of *Dnmt1* has been generated that exhibits a developmental arrest phenotype that is very similar to those observed for targeted deletion mutants (Takebayashi et al., 2007). This suggests that the catalytic function of DNMT1 is very important in early mouse embryogenesis, although undifferentiated embryonic stem cells are relatively unaffected by loss of DNMT1 activity (Jackson-Grusby et al., 2001; Tsumura et al., 2006). By contrast, complete inactivation of DNMT1 in human cancer cells leads to activation of a G2/M checkpoint and mitotic catastrophe with minimal changes in DNA methylation levels (Chen et al., 2007). Taken together these studies suggest the reported phenotypes of DNMT1 depletion in early development and cell lines may reflect the loss of both its enzymatic and non-enzymatic functions.

Although there is no evidence of global demethylation, imprinting, or inactivation of sex-specific chromosomes in *Xenopus laevis*, we have shown previously that xDnmt1 has an essential function in maintaining gene silencing prior to zygotic gene activation at the mid-blastula transition (MBT) in early amphibian development (Stancheva and Meehan, 2000). However, it was not clear from this study whether maintenance of gene silencing prior to the MBT in *Xenopus* depended on the enzymatic or non-enzymatic functions of xDnmt1. Here, we show that the silencing function of xDnmt1 in early amphibian development is independent of its methyltransferase activity. We report that a partial reduction in xDnmt1p levels by morpholino (xDMO) injection into *Xenopus laevis* embryos results in premature zygotic gene activation without a concomitant decrease in DNA methylation levels, either globally or at specific loci. Rescue experiments with an mRNA encoding a catalytically inactive form of human Dnmt1 (DNMT1) strongly suggest that DNA methylation is not used as a general silencer of gene expression in *Xenopus* embryos. Our data support a model in which xDnmt1 can regulate embryonic gene silencing directly and independently of its catalytic function.

MATERIALS AND METHODS

Oligonucleotides and morpholinos

The sequences of the primers and morpholinos used in this study are shown in Table 1.

Embryo manipulations

Xenopus embryos were handled as described (Ruzov et al., 2004). Morpholino oligonucleotides against xDnmt1 (xDMO) were designed and synthesised by GeneTools. Human rescue mRNAs

[1]Human Genetics Unit, MRC, Western General Hospital, Crewe Road, Edinburgh EH4 2XU, UK. [2]Genes and Development group, School of Biomedical Sciences, The University of Edinburgh, Hugh Robson Building, George Square, Edinburgh EH8 9XD, UK. [3]Institute of Gene Biology, Russian Academy of Sciences, Vavilova 34/5, Moscow, 119334, Russian Federation.

*Author for correspondence (e-mail: richard.meehan@hgu.mrc.ac.uk)

Accepted 16 January 2008

were synthesised from wild-type or mutant (C1226Y) DNMT1 plasmids (gift from Michael Rountree) using the T3/T7 Capscribe kit (Boehringer).

TNT
cDNA encoding full-length xDnmt1 was used as a template in coupled *in vitro* transcription-translation (TNT, Promega) reactions performed in the absence or presence of xDMO followed by PAGE.

RT-PCR
RT-PCR was performed as published (Ruzov et al., 2004).

In situ analysis
In situ protocols were performed using standard methods.

Southern blotting
Southern blotting was performed as described (Stancheva and Meehan, 2000). xSatellite I (xSatI) probe was generated by PCR.

Bisulfite sequencing
The bisulfite sequencing protocol has been described previously (Stancheva et al., 2002).

Promoter cloning
xOct91 and *xOct25* promoter regions were cloned from *Xenopus laevis* genomic DNA using the DNA Walking Kit (SeeGene, Korea). *xOct60* promoter was cloned by synteny PCR and the xOct91 promoter region (–463 to –12) was cloned into *Sac*I/*Bgl*II sites of pGL3-Luc basic.

Western blotting
Embryonic extracts were isolated using RIPA buffer and xDnmt1 levels were detected by immunoblotting with α-xDnmt1 antibody 3C6 (Shi et al., 2001). Mouse cell extracts were prepared and the following antibodies were used: α-T7 (T7-xSp1) (Novagen); α-human DNMT1 (NEB); and αPCNA (Abcam). Embryonic histone extracts were prepared by acid extraction and blotted with the following antisera: panAcH4 (Cell Signalling, 9441S), H3K9Ac (Abcam, AB4441), H4K5Ac (Upstate, 06-759), PanMethKH3 (Abcam, AB7315), H3K4me3 (Abcam, AB8580), H4K20me3 (Abcam, AB9053), H3K9me3 (Abcam, AB8898 and H3 (Abcam, AB1791).

GST pull down
Binding reactions for DNA GST pull downs were prepared as for EMSA (Ruzov et al., 2004). CpGpos oligonucleotide probes were used (Voo et al., 2000). Reactions were incubated for 10 minutes on a shaker at 30°C, washed four times with PBS, treated with Proteinase K, extracted and analysed using PAGE.

Transient transfections and reporter assays
Human 293T cells and mouse N2A cells were cultured using standard methods. Constructs for reporter assays were transfected into Neuro2A cells using established methods (Invitrogen). Assays were carried out independently in quadruplicate.

ChIP analysis
Xenopus A6 cells were transfected with xDnmt1-GFP, pCMVxDnmt1 or without plasmid DNA. CHIP was performed with a GFP antibody (Abcam).

RESULTS

xDnmt1p reduction causes premature gene activation

In previous work we used an antisense RNA (AS) strategy to knockdown xDnmt1p transiently in early embryos; it resulted in DNA hypomethylation and premature gene activation (Stancheva and Meehan, 2000). Here, we use highly stable (xDMO) morpholinos, which inhibit xDnmt1 mRNA translation in vitro and in vivo (Fig. 1A), and result in a phenotype that is indistinguishable from antisense RNA depletion. The xDMO morphants develop normally up to MBT but at gastrulation exhibit an extended open blastopore, which by neurulation results in the appearance of dead shedding white cells on the surface of a high proportion (80%) of embryos (arrows in Fig. 1B and insert) and a failure to form a neural tube. By tadpole stage, only 2% of the xDMO morphants are phenotypically normal compared with 91% for controls (see Fig. 3). This is intriguing as antisense injection results in almost complete depletion of xDnmt1 in stage 8 pre-MBT embryos (Stancheva and Meehan, 2000), whereas xDMO injection results in a 40-50% reduction of xDnmt1 protein (Fig. 1A). Despite this difference, the phenotypes of the AS and xDMO embryos are virtually indistinguishable.

Array screens in our laboratory (D.S.D. and R.R.M., unpublished) indicate that up to 25% of genes in these experiments are mis-expressed in stage 8 xDMO morphants. We used RT-PCR to verify the expression of these putative methyl-CpG dependent target genes in pre-MBT (stage 7-8) embryos (wild type and xDMO morphants). All tested genes were mis-expressed in xDMO morphants (*xCycD1*, *xSox17β*, *xMix1*, *xp68*, *xDep, xOct91* and *xID2*) relative to histone H4 and *xOct60* expression (Fig. 1C; see Fig. S1 in the supplementary material). Whole-mount RNA in situ hybridisation revealed that ectopic transcripts are present throughout the animal pole of xDMO stage 8 morphants (Fig. 1D). A control shows equal expression of the maternal oocyte-specific gene *xOct60* between the two embryo sets. We conclude that the xDnmt1p reduction in xDMO morphants is sufficient for premature gene activation before MBT, the induction of apoptosis and phenotypic defects that result in reduced survival rates. More importantly, premature gene activation is a general feature of xDMO embryos, implying an essential global role for xDnmt1p in embryonic gene repression.

xDMO morphants retain normal DNA methylation at repeat and unique sequences

To determine whether global methylation levels were altered in xDMO morphants, we tested the dispersed satellite I repeat (xSatI) by Southern blotting, which is methylated at its two *Hpa*II (C<u>CG</u>G) sites through development (Stancheva et al., 2002). Genomic DNA from both wild-type and xDMO siblings showed a comparable resistance to *Hpa*II digestion either by itself or in double digestion with *Hin*dIII (Fig. 2A). We used bisulphite sequencing analysis to precisely map CpG methylation at xSatI sequences in wild-type and xDMO genomic DNA. No hypomethylated CpGs were observed in xDMO DNA relative to the wild type (Fig. 2B; boxed numbers indicate % methylation at each CpG). As xSatI is distributed through the *Xenopus* genome, this suggests there are no genome-wide changes in DNA methylation in xDMO morphants.

Subsequent to finding normal methylation patterns in xDMO repeat DNA, the next issue was to evaluate the methylation profile of xDMO target genes (*xOct91* and *xCycD1*, Fig. 2C). This analysis showed that the pattern of methylation at *xOct91* and *xCycD1* promoters and upstream regions in stage 8 xDMO morphants was identical to stage 8 wild type (Fig. 2D). *xOct91* and *xCycD1* are zygotically activated during normal development after MBT (see Fig. S2A in the supplementary material) so we compared bisulfite maps when they are either transcriptionally silent (stage 8) or active (stage 10). There was no significant difference in CpG methylation at the *xOct91* and *xCycD1* loci between the inactive and active stages (data not shown), indicating that DNA methylation does not play a direct role in regulating their expression during normal development. We made similar observations for the *xOct25* and xSox17β promoter regions (data not shown). Together with the xSatI methylation analysis, this led us to conclude that DNA methylation is not directly regulating the expression of the *xOct91* and *xCycD1* loci in xDMO morphants and by extension other genes

that are misexpressed in stage 8 xDMO embryos. Our data imply that premature gene activation during *Xenopus* embryogenesis is governed by a mechanism independent of DNA modification.

Several studies have identified crosstalk between the DNMT1 proteins and histone modifying enzymes (Fuks, 2005). One possibility is that premature transcription in xDMO morphants may be due to global alterations in histone modification states as a consequence of xDnmt1p depletion. We tested this possibility by direct comparisons of histone mark abundance levels by immunoblotting. These experiments revealed no significant differences for various histone acetylation and histone methylation marks globally in early (stage 8) or later (stage 15) xDMO morphants (Fig. 2E). In stage 8 embryos, most histone marks are low to undetectable and only accrue as development proceeds, particularly in the case of H3K4me3 (see Fig. S2 in the supplementary material) and H4K20me3. However, these experiments cannot completely rule out subtle histone mark changes at specific gene promoters. Our data imply that neither global

Table 1. Primer and morpholino sequences used

Morpholino	Sequence
(xDMO) and xDnmt1a	GGACAGGCGTGAAACAGACTCGGC
xDnmt1b	GAACAGGCGTGAGACACACTCGGC
Control	CGCTCAGCTCCTCCATGTCTGCCGC
RT-PCR primer	**Sequence**
xId2u	ctcctgtacaatatgaatgattg
xId2l	acaagatgctgatgtctgtg
xp68u	ttgatgaagcagacagaatg
xp68l	cgttacacacatcaacaatc
xOct60u	ccatattgtacagccaaacctc
xOct60l	gttccagtcaaaggaagcag
xTrip7u	tcatcaaaacccgaacctc
xTrip7l	caattcattctatctccgcac
xCycd1u	atttcaagtgcgtccagaag
xCycd1l	ggaattgtcggtgtaaatgc
xSox17Bu	gtcatggtaggagagaac
xSox17Bl	tctgtttagccatcactgg
xMix1u	cctaatagttcctcacatc
xMix1l	ttgaagtggtagatacagg
xDepu	agagcggaatggcaactgt
xDepl	caacgtccacagcctcaga
xH4u	cgggataacattcagggt
xH4l	tccatggcggtaactgtc
xOct91u	cagatggcagcggacag
xOct91l	caactggttggcagaatcc
xOct25u	taatggagagatgcttgatg
xOct25l	ttctctatgttcgtcctcc
xBf2u	cgaagagaccgatatcgatg
xBf2l	ctgcaggatggacatggtg
xGapdhu	tgccattctcagccttaac*
xGapdhl	acggatttggctgtattgg*
xOdcu	gtcaatgatggagtgtatggatc*
xOdcl	tccattccgctctcctgagcac*
xPtenu	taccaggaggatggattcg
xPtenl	ggttgtggtcttcaaacgg
xMatu	gcttctgtgccagttacg
xMatl	tctgccagaatagtccgc
Promoter cloning primer	**Sequence**
xOct25tsp1	ccaccagcactgaccttgtcacgag
xOct25tsp2	gtccaagtccttgaggtgcagg
xOct25tsp3	cactctgccatccttgcaacgg
xOct91tsp1	ctgatgttcccctatacagctc
xOct91tsp2	gagctcttgctcaaatcctacc
xOct91tsp3	ctaagttgctcggtgttcgctc
xOct25-oct60_U	ccacaaggttacatggcagctcc
xOct25-oct60_L	tggggaaggaggtttggctgtac
Bisulfite data primer	**Sequence**
xOct25 Bis1	attgatgtataataaaggagg
xOct25 Bis2	tttgtatgttatatagagtg
xOct25 Bis3	gttttggtttattatttattagg
xOct25 Bis4	cacttcactttacaacaactacc
xOct25 Bis5	ctttacaacacttcactttac
xOct25 Bis6	cttaacaaatctttacaacac
xOct25 Bis7	gggttaggttgttttgagttg
xOct25 Bis8	agagatggggattagatttg
xOct25 Bis9	tttgagttgagggtaaatag

Bisulfite data primer	Sequence
xOct25 Bis10	cctaaaaccaccaacactaac
xOct25 Bis11	caatcaaatcaactaaaacc
xOct25 Bis12	taatatccaaaaacttacaatc
xOct91 Bis1	gatatttatgaagttttattg
xOct91 Bis2	ttatgaagttttattggtgag
xOct91 Bis3	gttttttttatatgttaatgag
xOct91 Bis4	acccaattaataaaacttac
xOct91 Bis5	ctccaaataacataattctc
xOct91 Bis6	tttcctctactccaaataac
xOct91 Bis7	gaggtgttttattggatatg
xOct91 Bis8	gttttattggatatgtatag
xOct91 Bis9	atagattaatagtttaaatag
xOct91 Bis10	ccctatacaactcttactc
xOct91 Bis11	tatacataatctaatattcc
xOct91 Bis12	atctattaattatacataatc
xOct60 Bis1	aaatttttaagggtagaggttg
xOct60 Bis2	ggtagaggttgtatttttttaaag
xOct60 Bis3	aaagagggggttggttttttg
xOct60 Bis4	ttaactaaaaatacccaataac
xOct60 Bis5	atccttttaactaaaaatacccc
xOct60 Bis6	tacccaaatcatccttttaac
xOct60 Bis7	aattggttaatgagagagaag
xOct60 Bis8	tgagagagaagattaattagtg
xOct60 Bis9	aagattaattagtgggtttatg
xOct60 Bis10	ccaactcttccaacccaaaacc
xOct60 Bis11	aaactaccaactcttccaaccc
xOct60 Bis12	aaacccaaactaccaactcttc
xCycd1 Bis1	ttagtttgatatttgtgtttg
xCycd1 Bis2	ggtttaatttaagttttatag
xCycd1 Bis3	tttattttattggttaattgtatg
xCycd1 Bis4	actaaacacccaacatacac
xCycd1 Bis5	aaacacccaacatacacc
xCycd1 Bis6	taaaataaaacttcaacctac
xCycd1 Bis7	ttatgaatggaggggggtg
xCycd1 Bis8	gaggggggtgggtgtag
xCycd1 Bis9	caccaacctccaacatcc
xCycd1 Bis10	cccaatacccaaaaaaacac
xSatI Bis1	gttaatattaatttgaggtttag
xSatI Bis2	gtttgaatagtttagttggtag
xSatI Bis3	aaatactaaataaaaaaaccc
xSatI Bis4	ttcaaactaatactaaacaaac
Miscellaneous primers	**Sequence**
xSat1u	catttgagaagctgactagc
xSat1l	agaaccactgtcctcattaatc
ChIP primers	**Sequence**
O25CHIP1	ccgcagttggaggggtggc
O25CHIP2	gtctcaaagaggccaatgaatg
O25CHIP3	ttgaggtgcaggaaagcaac
O25CHIP4	gtaatcactctgccatccttg
CycD1CHIP5	caagtggaagaagagctggc
CycD1CHIP6	gtgggtgcagaggctccttc
CycD1CHIP7	gccctcctgatggtgtccac
CycD1CHIP8	gatgaggttcctgtccagatg
xlTubChipU	tgaaacaggagcaggaaagc
xlTubChipL	gctctgggtggaataactgc

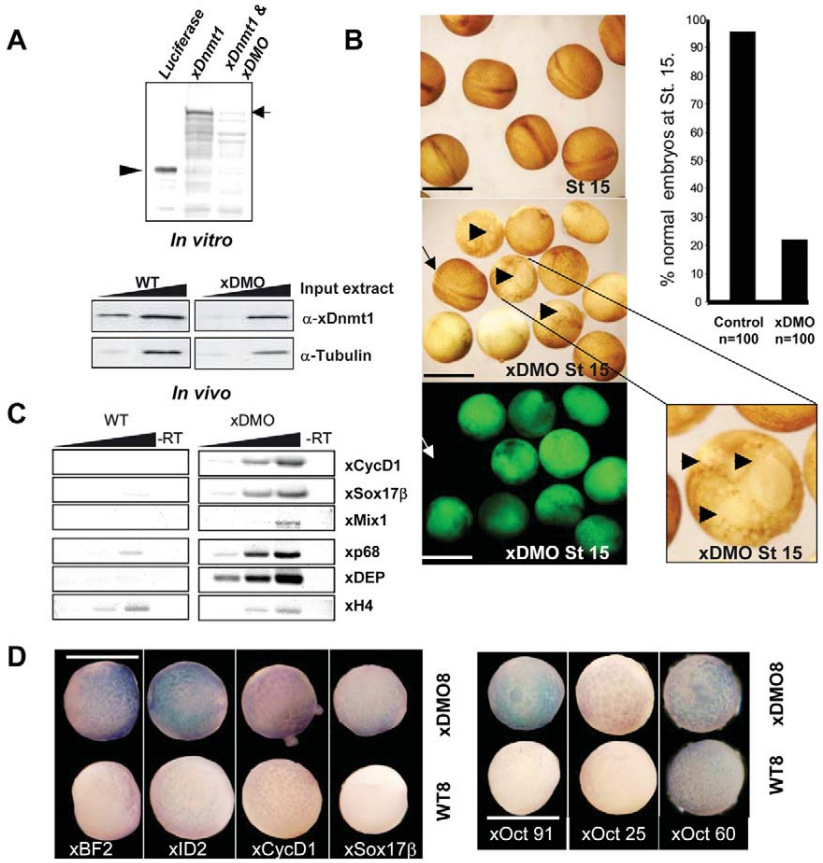


Fig. 1. xDMO embryos have reduced xDnmt1 levels, are abnormal and mis-express genes. (**A**) Top panel: in vitro inhibition of xDnmt1 translation (black arrowhead) using xDMO (compare lanes 2 and 3). Bottom panel: in vivo inhibition of xDnmt1 translation in pre-MBT (stage 7-8) embryos (compare wild-type and xDMO extracts). Tubulin is used as a loading control. (**B**) Left panel: phenotypes of stage 15 embryos. Morphant xDMO embryos exhibit apoptotic lesions (arrowheads and enlargement) and lack neural folds (black arrow) compared with control stage 15 embryos. xDMO embryos contain fluorescein, unlike the control embryo (compare arrowed embryos). Right panel: comparison of percentage (*n*=100) of successfully neurulating embryos for wild type and xDMO. (**C**) xDMO embryos mis-express a range of transcripts. Wild-type and xDMO RNA was assayed by RT-PCR over a 10-fold dilution range (0.1, 0.3 and 1 μl cDNA for each sample indicated by the black triangles). H4 is a loading control. (**D**) In situ analysis reveals ectopic expression of the indicated xDMO targets throughout the animal pole (compare wild-type and xDMO panels). The maternally expressed gene *xOct60* is not mis-expressed. Scale bars: 1 mm in B,D. Animal pole views are shown.

changes in DNA methylation or specific histone modifications are associated with activation of xDMO target genes in stage 8 embryos, but rather that high levels of xDnmt1p that are present in early *Xenopus* embryos are essential for their repression (Shi et al., 2001).

Dnmt1p catalytic activity is not essential for repression

To explore the possibility that xDnmt1 may have a non-enzymatic role in gene silencing, we carried out a series of rescue experiments with wild-type and mutant forms of human DNMT1 (Fig. 3A). Two-cell embryos were injected with xDMO alone or in combination with either wild type or mutant human DNMT1 (hDNMT1^{C1226Y}) mRNA (which xDMO does not bind), and fixed for whole-mount RNA in situ analysis. *xBF2* and *xOct25* were ectopically activated (in agreement with our array and RT-PCR screens) in xDMO morphants, but the presence of either wild-type DNMT1 or hDNMT1^{C1226Y} mRNA (1 ng) significantly reduced the extent of activation by up to 50% as measured by densitometry (Fig. 3B; see Fig. S3B in the supplementary material). In two further series of experiments, both types of hDNMT1 mRNAs increased the frequency of phenotypically normal embryos at stage 15 (neurulation) two-fold (from 20% to more than 40%), indicating the specificity of the morpholino, but more importantly that the catalytic function of human DNMT1 is not required to rescue the xDMO morphant phenotype and re-impose gene silencing (Fig. 3C,D). A high proportion of rescued embryos do not show evidence of developmental delay and form neural folds (Fig. 3D, black arrows) equivalent to those seen in wild-type neurula embryos. Unlike xDMO morphants, the rescued embryos continue to develop and can form tadpoles at a high frequency (Fig. 3E). The major conclusion from these developmental studies is that a catalytically inactive human mRNA can partially restore the normal transcriptional program and phenotype in *Xenopus* embryos, thereby underlining a physiologically relevant non-enzymatic role for xDnmt1p in gene repression.

xDnmt1p is a transcriptional repressor and can be localised to target gene promoters

In light of the above data, we sought to explore potential mechanisms for DNMT1-mediated repression. The N-terminal non-catalytic region of mammalian DNMT1 is an effective transcription repressor when artificially recruited to a promoter (Fuks, 2005), and knockdown of DNMT1 in transformed cells specifically activates expression of two genes, independently of DNA methylation (Milutinovic et al., 2004). We hypothesised that if xDnmt1, like its mammalian counterparts, contained regions that can directly bind non-methylated DNA, this would enable it to act as a general repressor of transcription during early *Xenopus* development (Chuang et al., 1996; Suetake et al., 2006). We tested three candidate regions (Fig. 4A, black bars G1-G3) of xDnmt1 as GST fusion proteins for DNA binding activity in vitro using a pull down assay. Under the stringent conditions employed, all three GST fusion proteins bound the CpGpos oligonucleotide (Voo et al., 2000) as shown in Fig. 4A. Additional experiments suggest that xDnmt1 binding has relaxed sequence specificity (data not shown).

We tested whether xDnmt1 can repress the activation of a minimum promoter that has four copies of the xSp1-binding site driving luciferase expression (p4xSp1-Lucif) (Kockar et al., 2001). Relative induction by xSp1 is reduced by 55% in the presence of

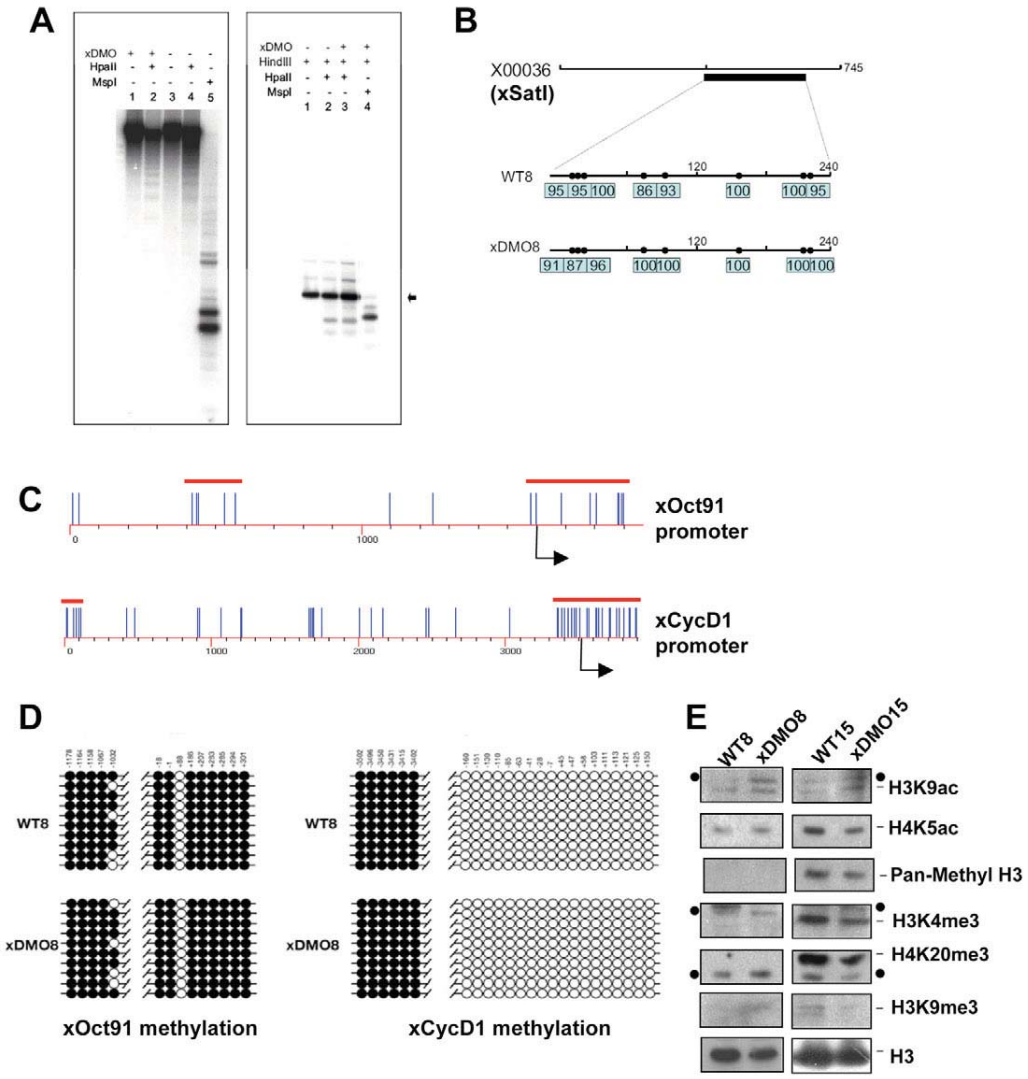


Fig. 2. No changes in DNA methylation at repeat or single copy sequences in xDMO morphants. (**A**) xDMO DNA is heavily methylated at xSatI *Hpa*II sites (compare lanes 2 and 4). *Hpa*II is a methyl-sensitive restriction enzyme and *Msp*I is the methylation insensitive (CCGG) counterpart (lane 5). Right panel: *Hin*dIII was used to generate the 750 bp xSatI monomer (black arrow); double digestion with *Hin*dIII and *Hpa*II showed no difference in monomer methylation in the wild-type and xDMO samples (lanes 2-3). (**B**) Bisulphite sequencing (clones *n*=40) shows no significant difference in CpG methylation between wild-type and xDMO genomes at xSatI sequences. Boxed numbers are percentage CpG methylation; black circles indicate CpG distribution in xSatI. (**C**) CpG distribution in cloned promoters of *xOct91* and *xCycD1*. Blue bars, CpG; black arrows, transcription start sites; red bars, regions sequenced. (**D**) Bisulfite analysis (sequences *n*=40, ten representative clones are shown) was used to determine the methylation status of *xOct91* (left) and *xCycD1* (right) promoters and upstream regions. Numbers above each CpG indicate genomic position relative to transcription start. Filled circles, methylated CpGs; empty circles, non-methylated CpGs. (**E**) Immunoblot analysis of wild-type and xDMO histones shows no significant change in various histone modification marks between histone WT and xDMO extracts at stages 8 and 15. Histone modifications are low to absent at stage 8 and accrue by stage 15. Black dots indicate non-specific bands.

xDnmt1 and both the catalytically active and inactive forms of human DNMT1 (Fig. 4B), which are expressed equally in transient transfection assays (Fig. 4C, left panel). Analysis of the raw luciferase data suggests that as the Dnmt1 dose is increased, p4xSp1 activation is repressed up to 10-fold without affecting cell numbers (data not shown). However, the expression of the co-transfected luciferase (*Renilla*) reporter is concomitantly repressed by fivefold and results in a normalised value of a twofold (50%) reduction. Western blot analysis shows that there is no observable difference in xSp1 levels between cells transfected with a 10-fold difference of the xDnmt1 expression plasmid (Fig. 4C, right panel). Our data suggest that untethered DNMT1 can act as a general repressor of promoters when it is abundant, and that its catalytic activity is dispensable for this function.

To reflect the *xOct91* repression scenario in early development, we measured the effect of Dnmt1 overexpression on *xOct91* promoter activity. Each form of DNMT1 repressed the activity of the *xOct91* reporter construct by ~50% or more relative to the empty vector control (Fig. 4D). The repressive capacity of hDNMT1^{C1226Y} strongly indicates that the mechanism of inhibition of these transgenes is independent of DNA methylation. In addition, repression by DNMT1s is not relieved by an HDAC inhibitor (TSA) and requires full-length

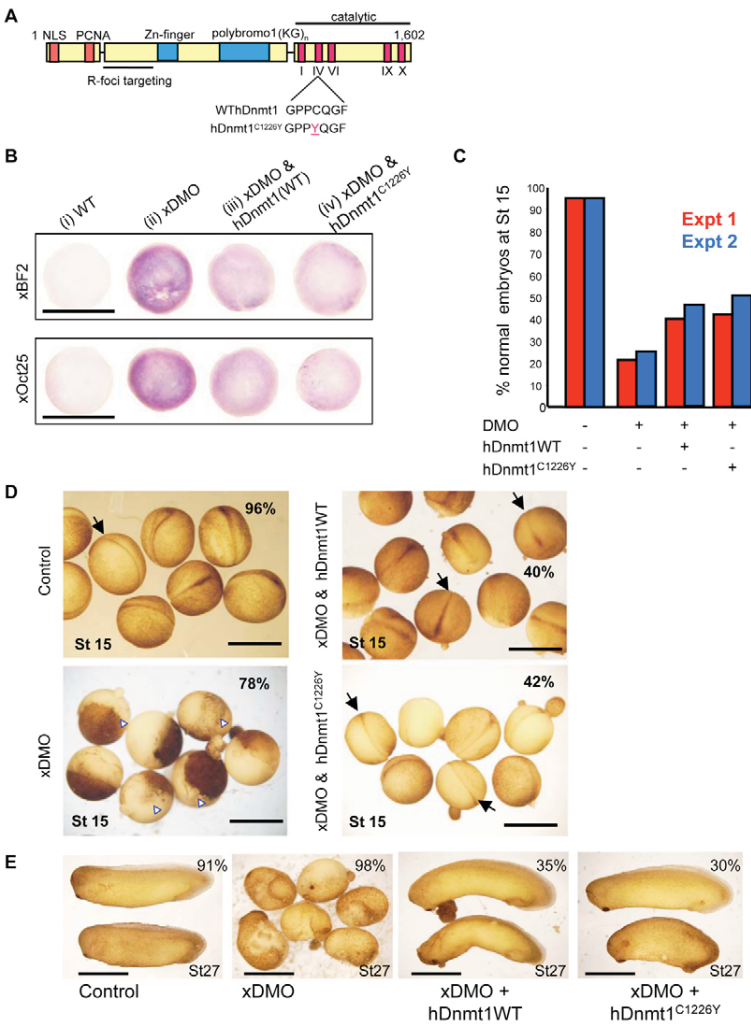


Fig. 3. A catalytically inactive form of human DNMT1 can restore repression and rescue xDMO morphants. (**A**) Functional domains present in human DNMT1 and an inactivating point mutation of hDNMT1 (hDNMT1^{C1226Y}) (Jair et al., 2006). (**B**) Catalytic function of human DNMT1 is not required to rescue *xOct25* and *xBF2* gene repression. Compare rescue in situ intensities (iii and iv) with signals in control (i) and xDMO morphants (ii). Animal pole views are shown. (**C**) hDNMT1WT and hDNMT1^{C1226Y} restore normal *Xenopus* neurulation with comparable efficiency. Experiment 1 (n=100) is in red and experiment 2 (n=60) is in blue. Compare number of normally neurulating embryos in xDMO only (~20%) with rescued injected embryos (>40%). (**D**) Rescued embryos are similar in phenotype to wild-type siblings. Neurulating embryos are shown. There are apoptotic cells and open blastopores (white arrowheads) in xDMO embryos (bottom left). Such lesions are absent in rescued embryos (right panels), which are similar to control injected embryos (top left) (black arrows indicate neural folds). (**E**) Phenotypes of late stage (stage 27) embryos. Only 2% (i.e. 98% of embryos fail) of xDMO morphants develop to late tadpole stage (stage 27) compared with 35% and 30% for wild-type (hDNMT1WT) and mutant (hDNMT1^{C1226Y}) rescued embryos, which develop normally. Scale bars: 1 mm.

forms of the protein for efficient silencing and xDMO rescue (J.A.H., A.R. and R.R.M., unpublished). We conclude that DNMT1 can act as a general repressor of non-methylated promoters and that its catalytic activity is not required for this function.

As a putative repressor of gene expression and possessing DNA affinity, we hypothesised that xDnmt1 should bind the loci it regulates. We were unable to perform ChIP assays against endogenous xDnmt1 using monoclonal antibodies as they were sensitive to formaldehyde crosslinking (D.S.D., unpublished). We performed ChIP with *Xenopus laevis* A6 cells that were transfected with a GFP-xDnmt1 expression construct and an anti-GFP polyclonal antibody. This analysis localised GFP-xDnmt1 to the *xOct25* and *xCycD1* non-methylated promoters in A6 cells (Fig. 4E,F) but not to an intronic region of the constitutively active *Xenopus* α-Tubulin gene. Enrichment of GFP-xDnmt1 at non-methylated, non-expressed loci (xDMO target genes) is consistent with a model in which xDnmt1p can bind and repress gene expression independently of DNA methylation during development.

DISCUSSION

Our experiments and recent reports both suggest that DNMT1 has an essential role in early amphibian and mouse development (Takebayashi et al., 2007). Mutant mice expressing a DNMT1 point-mutant protein lacking catalytic activity (DNMT1-C1229S) fail to develop, arrest after gastrulation (E9.5) with a near-complete loss of DNA methylation and mis-express normally methylated genes with phenotypes very similar to those of the *dnmt1$^{c/c}$* mutant (Lei et al., 1996; Takebayashi et al., 2007). This indicates that the catalytic function of *Dnmt1* is required to support early mouse embryogenesis, probably owing to the many developmental processes, including X-chromosome inactivation, suppression of retrotransposon activity, imprinting and regulation of germ-cell specific gene expression, that are dependent on DNA methylation (Goll and Bestor, 2005; Maatouk et al., 2006). By contrast, *Xenopus laevis* lacks imprinting and processes equivalent to X-inactivation, which may underlie its non-dependence on DNA methylation in early development. Additionally primordial germ cell development is specified by different mechanisms in amphibians and mammals (Crother et al., 2007). Our work suggests that the intrinsic repression function of xDnmt1p has been used to maintain gene silencing in pre-MBT embryos until co-ordinated and precise zygotic activation occurs at multiple gene loci. This period of gene silencing over 11 cell divisions is not observed during mouse embryogenesis (Meehan et al., 2005) and implies distinct regulatory mechanisms for zygotic gene activation are used between mice and frogs.

Zygotic activation of gene expression during early *Xenopus* development is mainly dictated by maternal inheritance of repressor and activator components (Veenstra, 2002). A two-cell *Xenopus*

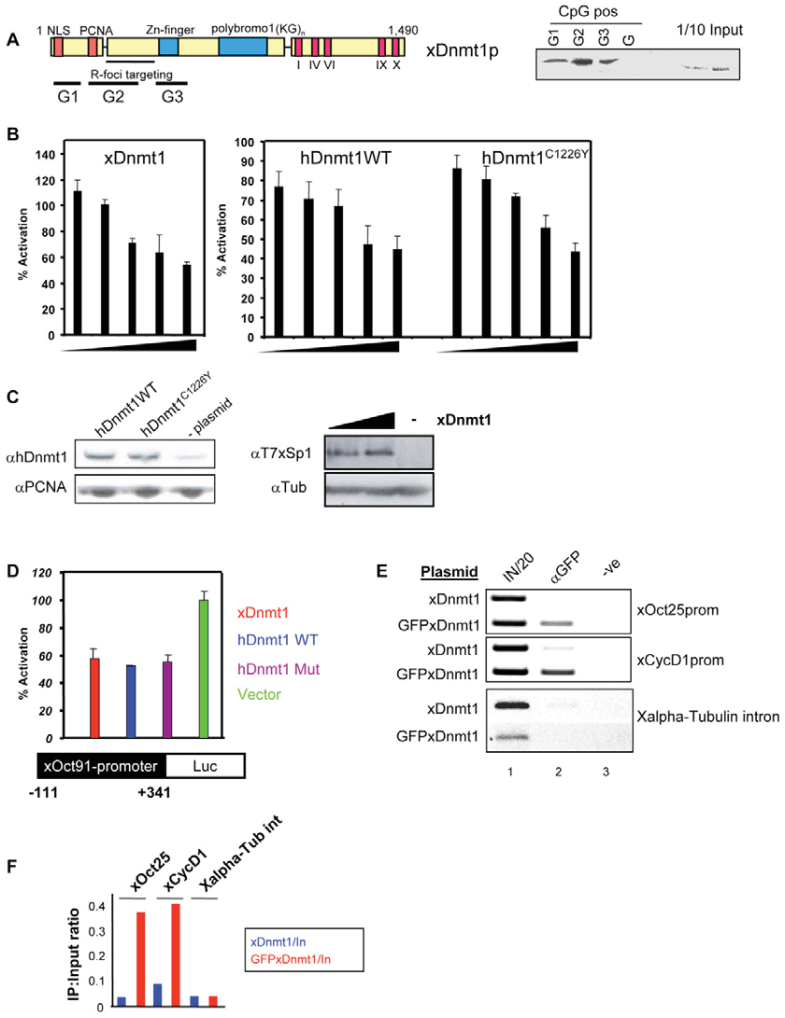


Fig. 4. Dnmt1 represses activation of non-methylated transgenes in vivo and xDnmt1 localises to non-methylated target genes. (**A**) N-terminal xDnmt1 fusions bind dsDNA oligos. The domain structure of xDnmt1 (top). xDnmt1-Gst fusions domains G1-G3 (black bars). CpGpos oligonucleotides were used in pull-down assays with the three Gst fusions and Gst only. All three constructs bound CpGpos compared with the Gst protein (G) lane. (**B**) In vivo assays to test the repression activities of xDnmt1, hDNMT1WT and hDNMT1^{C1226Y}. Exogenous xSp1 was used to activate the Sp1-Luc reporter in the presence of increasing amounts of Dnmt1 [black triangles; 0.125-2 μg plasmid DNA per transfection]. xSp1 activation of the reporter alone was assigned 100%. Data were obtained from nine independent assays and normalised to TK-Renilla. (**C**) Left panel: both forms of human DNMT1 (hDNMT1WT and hDNMT1^{C1226Y}) are expressed equally after transfection into 293T cells relative to PCNA and endogenous hDNMT1 (–plasmid). Right panel: N2A cells were transfected with T7xSp1 against low or high xDnmt1 levels (black bar), cell extracts were blotted with α-T7. αTubulin was used as a loading control. (**D**) An *xOct91* promoter (–111 to +343) reporter construct is repressed by co-transfection with 500 ng of xDnmt1, hDNMT1WT and hDNMT1^{C1226Y} but not by the empty vector control. (**E**) Chromatin IP (ChIP) analysis shows recruitment of GFP-xDnmt1 to the non-methylated *xOct25* and *xCycD1* promoters, but not an α-tubulin intron in A6 cells. Note the enrichment of GFP-xDnmt1 at both promoters using α-GFP (lane 2) but not the control xDnmt1 lacking GFP (panel xDnmt1). Lane 1, input (1/20 used in IP); –ve, no antibody control. (**F**) Bar chart shows fold enrichment of GFP-xDnmt1 at the *xOct25* (eightfold) and *xCycD1* (4.5-fold) promoters compared with a non-tagged control and with αTubulin.

embryo contains up to 100 ng of histones and 10 ng of xDnmt1 (Shi et al., 2001; Veenstra, 2002), which together impose dominant repression of transcription. At MBT, virtually all of the free histone pool has been depleted and we hypothesize that transcription repression is now sensitive to xDnmt1 protein levels. Radioactive labelling experiments and successful depletion by antisense RNA and morpholinos suggest that continuous translation is required to maintain high xDnmt1p levels in oocytes and embryos (Kimura et al., 1999). The rate of transcription per cell increases ~200-fold when xDnmt1p levels are reduced to approximately 2 pg/cell in normal embryos (Newport and Kirschner, 1982a; Hashimoto et al., 2003). In xDMO morphants, a 40-50% reduction in xDnmt1p before the MBT is sufficient to allow the low level of general transcription machinery components that are present, such as RNA Pol I, Pol II and TBP, to dramatically activate a generalised pattern of gene expression approximately two cell cycles earlier than normal. However, this pattern of gene activation does not equate with a normal transcriptional activation profile as the spectrum of genes that are mis-expressed in xDMO morphants (and antisense-depleted embryos) is much greater, as measured by cDNA library screens and array analysis (D.S.D. and R.R.M., unpublished). This suggests that the transcriptional competence of normal late blastula/early gastrula embryos is crucially dependent on the non-catalytic and catalytic silencing function of xDnmt1. In general, transcription at MBT occurs before histone modifications, such as H3K4M3 (Fig. 2E) or H4 acetylation, accumulate (Almouzni et al., 1994). This may explain why we do not observe an accumulation of histone modifications in pre-MBT xDMO morphants. It is possible that changes in histone modifications during *Xenopus* development are linked with a dynamic alteration in the organization of different chromatin domains that occurs after the MBT when gene-specific subdomains are set up (Vassetzky et al., 2000).

By depleting xDnmt1p in stage 8 embryos, we propose that this interferes with its non-catalytic repression function, which is crucially dependent on its high abundance in early *Xenopus* embryos. Some developmentally decisive genes, such as *xBra*, remain silenced unless DNA hypomethylation also occurs (see Fig. S4 in the supplementary material). In this instance, we suggest that *xBra* must represent a minor class of genes that are regulated directly by DNA methylation, as the extent of transcription activation in antisense and xDMO stage 8 embryos is essentially equivalent (data not shown). A generalised interpretation of the data is that repression by xDnmt1 in vivo is bimodal. In early development, xDnmt1 functions as a maintenance methyltransferase to perpetuate patterns of DNA methylation globally and at distinct loci (*xBra*). At these loci, silencing is dependent on the catalytic function of Dnmt1 and potential interpretation of the methyl-CpG mark. However, our data suggest that xDnmt1p may also act as a direct titratable repressor component at multiple loci that has been previously hypothesised to be present in *Xenopus* embryos (Newport and Kirschner, 1982b).

It is possible that mammalian DNMT1 also has multiple silencing functions because screens of mis-expressed genes in *Dnmt1*$^{-/-}$, *Trp53*$^{-/-}$ MEFs indicate that a high proportion (up to 80%) of mis-expressed genes have CpG island promoters, which would be predicted to be methylation free at all times (Jackson-Grusby et al., 2001; Lande-Diner et al., 2007). Similar to our xDMO targets, it is highly possible that these genes are inhibited through direct action of the Dnmt1 protein at promoters, hinting at conservation of non-methyl dependent functions. Recent work demonstrates that complete inactivation of DNMT1 function in human cancer cells results in cell death (Chen et al., 2007), but this decrease in viability occurs with minimal changes in global DNA methylation. This observation supports the hypothesis that DNMT1 possesses essential functions independent of its role as a maintenance methyltransferase, and links its absence with activation of a cellular checkpoint response. Unlike the situation in tumour cells (Spada et al., 2007), limited reduction in xDnmt1p levels appears to be sufficient to activate a cell death program in *Xenopus* embryos. The pathway that mediates activation of apoptosis in DNMT1-deficient cells is dependent on TRP53 function, but the activating signal has yet to be identified (Jackson-Grusby et al., 2001; Stancheva et al., 2001). It has also been observed that undifferentiated *Dnmt1*$^{c/c}$ ES cells have a growth advantage compared with wild-type controls, but this advantage was lost in the presence of a normal or mutant (C1229S) DNMT1 mini-genes (Damelin and Bestor, 2007). These observations support multifunctional non-enzymatic roles for DNMT1 in development, cellular differentiation and cancer.

We thank Sari Pennings, Wendy Bickmore and Hazel Cruickshanks for helpful comments and corrections during manuscript preparation, and Nick Hastie for general support. We thank Irina Stancheva for advice during early stages of this work. This work was first supported by Wellcome Trust project grants to R.R.M. Current work in R.R.M.'s laboratory (R.R.M., D.S.D., A.R. and J.A.H.) is supported by the MRC. We thank Mike Rountree, Adrian Bird, Richard Harland, Hugh Woodland and Dipak Ramji for plasmids. We thank Shoji Tajima for aliquots of the xDnmt1 monoclonal antibodies. We thank members of the Chromosome and Gene Expression Group, especially Jeremy Sanford, for technical advice. This paper is dedicated to the memory of our late colleague, John W. Newport, a trailblazer of the MBT.

Supplementary material

Supplementary material for this article is available at http://dev.biologists.org/cgi/content/full/135/7/1295/DC1

References

Almouzni, G., Khochbin, S., Dimitrov, S. and Wolffe, A. P. (1994). Histone acetylation influences both gene expression and development of Xenopus laevis. *Dev. Biol.* **165**, 654-669.

Chen, T., Hevi, S., Gay, F., Tsujimoto, N., He, T., Zhang, B., Ueda, Y. and Li, E. (2007). Complete inactivation of DNMT1 leads to mitotic catastrophe in human cancer cells. *Nat. Genet.* **39**, 391-396.

Chuang, L. S., Ng, H. H., Chia, J. N. and Li, B. F. (1996). Characterisation of independent DNA and multiple Zn-binding domains at the N terminus of human DNA-(cytosine-5) methyltransferase: modulating the property of a DNA-binding domain by contiguous Zn-binding motifs. *J. Mol. Biol.* **257**, 935-948.

Crother, B. I., White, M. E. and Johnson, A. D. (2007). Inferring developmental constraint and constraint release: primordial germ cell determination mechanisms as examples. *J. Theor. Biol.* **248**, 322-330.

Damelin, M. and Bestor, T. H. (2007). Biological functions of DNA methyltransferase 1 require its methyltransferase activity. *Mol. Cell. Biol.* **27**, 3891-3899.

Fuks, F. (2005). DNA methylation and histone modifications: teaming up to silence genes. *Curr. Opin. Genet. Dev.* **15**, 490-495.

Goll, M. G. and Bestor, T. H. (2005). Eukaryotic cytosine methyltransferases. *Annu. Rev. Biochem.* **74**, 481-514.

Hashimoto, H., Suetake, I. and Tajima, S. (2003). Monoclonal antibody against dnmt1 arrests the cell division of xenopus early-stage embryos. *Exp. Cell Res.* **286**, 252-262.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E. et al. (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. *Nat. Genet.* **27**, 31-39.

Jair, K. W., Bachman, K. E., Suzuki, H., Ting, A. H., Rhee, I., Yen, R. W., Baylin, S. B. and Schuebel, K. E. (2006). De novo CpG island methylation in human cancer cells. *Cancer Res.* **66**, 682-692.

Kimura, H., Suetake, I. and Tajima, S. (1999). Xenopus maintenance-type DNA methyltransferase is accumulated and translocated into germinal vesicles of oocytes. *J. Biochem.* **125**, 1175-1182.

Klose, R. J. and Bird, A. P. (2006). Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* **31**, 89-97.

Kockar, F. T., Foka, P., Hughes, T. R., Kousteni, S. and Ramji, D. P. (2001). Analysis of the Xenopus laevis CCAAT-enhancer binding protein alpha gene promoter demonstrates species-specific differences in the mechanisms for both auto-activation and regulation by Sp1. *Nucleic Acids Res.* **29**, 362-372.

Lande-Diner, L., Zhang, J., Ben Porath, I., Amariglio, N., Keshet, I., Hecht, M., Azuara, V., Fisher, A. G., Rechavi, G. and Cedar, H. (2007). Role of DNA methylation in stable gene repression. *J. Biol. Chem.* **282**, 12194-12200.

Lei, H., Oh, S. P., Okano, M., Juttermann, R., Goss, K. A., Jaenisch, R. and Li, E. (1996). De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells. *Development* **122**, 3195-3205.

Li, E., Bestor, T. H. and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-926.

Maatouk, D. M., Kellam, L. D., Mann, M. R., Lei, H., Li, E., Bartolomei, M. S. and Resnick, J. L. (2006). DNA methylation is a primary mechanism for silencing postmigratory primordial germ cell genes in both germ cell and somatic cell lineages. *Development* **133**, 3411-3418.

Meehan, R. R., Dunican, D. S., Ruzov, A. and Pennings, S. (2005). Epigenetic silencing in embryogenesis. *Exp. Cell Res.* **309**, 241-249.

Milutinovic, S., Brown, S. E., Zhuang, Q. and Szyf, M. (2004). DNA methyltransferase 1 knock down induces gene expression by a mechanism independent of DNA methylation and histone deacetylation. *J. Biol. Chem.* **279**, 27915-27927.

Newport, J. and Kirschner, M. (1982a). A major developmental transition in early Xenopus embryos: I. Characterization and timing of cellular changes at the midblastula stage. *Cell* **30**, 675-686.

Newport, J. and Kirschner, M. (1982b). A major developmental transition in early Xenopus embryos: II. Control of the onset of transcription. *Cell* **30**, 687-696.

Rai, K., Nadauld, L. D., Chidester, S., Manos, E. J., James, S. R., Karpf, A. R., Cairns, B. R. and Jones, D. A. (2006). Zebra fish Dnmt1 and Suv39h1 regulate organ-specific terminal differentiation during development. *Mol. Cell. Biol.* **26**, 7077-7085.

Ruzov, A., Dunican, D. S., Prokhortchouk, A., Pennings, S., Stancheva, I., Prokhortchouk, E. and Meehan, R. R. (2004). Kaiso is a genome-wide repressor of transcription that is essential for amphibian development. *Development* **131**, 6185-6194.

Shi, L., Suetake, I., Kawakami, T., Aimoto, S. and Tajima, S. (2001). Xenopus eggs express an identical DNA methyltransferase, Dnmt1, to somatic cells. *J. Biochem.* **130**, 359-366.

Spada, F., Haemmer, A., Kuch, D., Rothbauer, U., Schermelleh, L., Kremmer, E., Carell, T., Langst, G. and Leonhardt, H. (2007). DNMT1 but not its interaction with the replication machinery is required for maintenance of DNA methylation in human cells. *J. Cell Biol.* **176**, 565-571.

Stancheva, I. and Meehan, R. R. (2000). Transient depletion of xDnmt1 leads to premature gene activation in Xenopus embryos. *Genes Dev.* **14**, 313-327.

Stancheva, I., Hensey, C. and Meehan, R. R. (2001). Loss of the maintenance methyltransferase, xDnmt1, induces apoptosis in Xenopus embryos. *EMBO J.* **20**, 1963-1973.

Stancheva, I., El Maarri, O., Walter, J., Niveleau, A. and Meehan, R. R. (2002). DNA methylation at promoter regions regulates the timing of gene activation in Xenopus laevis embryos. *Dev. Biol.* **243**, 155-165.

Suetake, I., Hayata, D. and Tajima, S. (2006). The amino-terminus of mouse DNA methyltransferase 1 forms an independent domain and binds to DNA with the sequence involving PCNA binding motif. *J. Biochem.* **140**, 763-776.

Takebayashi, S. I., Tamura, T., Matsuoka, C. and Okano, M. (2007). Major and essential role for DNA methylation mark in mouse embryogenesis and stable association of DNMT1 with newly replicated regions. *Mol. Cell. Biol.* **27**, 8243-8258.

Tsumura, A., Hayakawa, T., Kumaki, Y., Takebayashi, S., Sakaue, M., Matsuoka, C., Shimotohno, K., Ishikawa, F., Li, E., Ueda, H. R. et al. (2006). Maintenance of self-renewal ability of mouse embryonic stem cells in the absence of DNA methyltransferases Dnmt1, Dnmt3a and Dnmt3b. *Genes Cells* **11**, 805-814.

Vassetzky, Y., Hair, A. and Mechali, M. (2000). Rearrangement of chromatin domains during development in Xenopus. *Genes Dev.* **14**, 1541-1552.

Veenstra, G. J. (2002). Early embryonic gene transcription in *Xenopus*. *Adv. Dev. Biol. Biochem.* **12**, 85-105.

Voo, K. S., Carlone, D. L., Jacobsen, B. M., Flodin, A. and Skalnik, D. G. (2000). Cloning of a mammalian transcriptional activator that binds unmethylated CpG motifs and shares a CXXC domain with DNA methyltransferase, human trithorax, and methyl-CpG binding domain protein 1. *Mol. Cell. Biol.* **20**, 2108-2121.

Walsh, C. P. and Bestor, T. H. (1999). Cytosine methylation and mammalian development. *Genes Dev.* **13**, 26-34.